

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083898 A1

- (51) International Patent Classification⁷: **C12N 15/12**, C07K 14/47, 16/18, C12N 15/63, 5/10, A01K 67/00, A61K 31/7088, 38/16, C12Q 1/68, G01N 33/53
- (21) International Application Number: PCT/IB01/00914
- (22) International Filing Date: 18 April 2001 (18.04.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (71) Applicant (*for all designated States except US*): **GENSET** [FR/FR]; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **BEJANIN, Stephane** [FR/FR]; 35, bd de Rochechouart, F-75009 Paris (FR). **TANAKA, Hiroaki** [FR/FR]; 8, av. de la Providence, F-92160 Antony (FR). **DUMAS MILNE EDWARDS, Jean-Baptiste** [FR/FR]; 8, rue Gregoire de Tours, F-75006 Paris (FR). **JOBERT, Severin** [FR/FR]; 7, impasse Tourneux, F-75010 Paris (FR). **GIORDANO, Jean-Yves** [FR/FR]; 12, rue Duhesme, F-75018 Paris (FR).
- (74) Common Representative: **GENSET**; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— with international search report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS

(57) Abstract: The invention concerns GENSET polynucleotides and polypeptides. Such GENSET products may be used as reagents in forensic analyses, as chromosome markers, as tissue/cell/organelle-specific markers, in the production of expression vectors. In addition, they may be used in screening and diagnosis assays for abnormal GENSET expression and/or biological activity and for screening compounds that may be used in the treatment of GENSET-related disorders.



WO 02/083898 A1

**FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED
PROTEINS**

CROSS REFERENCE TO RELATED APPLICATIONS

5 The present application is related to and claims priority from US provisional Application Nos. 60/197,873, filed April 18, 2000, 60/224,009, filed August 7, 2000, 60/260,328, filed January 8, 2001, and 60/224,006, filed August 4, 2000, the entire disclosures of each of which is herein incorporated by reference.

FIELD OF THE INVENTION

10 The present invention is directed to GENSET polypeptides, fragments thereof, and the regulatory regions located in the 5'- and 3'-ends of the genes encoding the polypeptides. The invention also concerns polypeptides encoded by GENSET genes and fragments thereof. The present invention also relates to recombinant vectors including the polynucleotides of the present
15 invention, particularly recombinant vectors comprising a GENSET gene regulatory region or a sequence encoding a GENSET polypeptide, and to host cells containing the polynucleotides of the invention, as well as to methods of making such vectors and host cells. The present invention further relates to the use of these recombinant vectors and host cells in the production of the polypeptides of the invention. The invention further relates to antibodies that specifically bind to
20 the polypeptides of the invention and to methods for producing such antibodies and fragments thereof. The invention also provides methods of detecting the presence of the polynucleotides and polypeptides of the present invention in a sample, methods of diagnosis and screening of abnormal GENSET polypeptide expression and/or biological activity, methods of screening compounds for their ability to modulate the activity or expression of the GENSET polypeptides, and uses of such
25 compounds.

BACKGROUND OF THE INVENTION

 The estimated 30,000-12,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable of specifically hybridizing to loci distributed throughout the human genome find
30 application in the construction of high resolution chromosome maps and in the identification of individuals.

 In the past, the characterization of even a single human gene was a painstaking process, requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and computer technology have merged to greatly accelerate the rate at which human genes can be
35 isolated, sequenced, mapped, and characterized.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed along the human genome. In one approach, large fragments of genomic DNA are isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bio-informatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bio-informatics software may mischaracterize the genomic sequences obtained, i.e., labeling non-coding DNA as coding DNA and vice versa.

An alternative approach takes a more direct route to identifying and characterizing human genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger RNAs (mRNAs) which encode human proteins. Using this approach, sequencing is only performed on DNA which is derived from protein coding fragments of the genome. Often, only short stretches of the cDNAs are sequenced to obtain sequences called expressed sequence tags (ESTs). The ESTs may then be used to isolate or purify cDNAs which include sequences adjacent to the EST sequences. The cDNAs may contain all of the sequence of the EST which was used to obtain them or only a fragment of the sequence of the EST which was used to obtain them. In addition, the cDNAs may contain the full coding sequence of the gene from which the EST was derived or, alternatively, the cDNAs may include fragments of the coding sequence of the gene from which the EST was derived. It will be appreciated that there may be several cDNAs which include the EST sequence as a result of alternate splicing or the activity of alternative promoters.

In the past, these short EST sequences were often obtained from oligo-dT primed cDNA libraries. Accordingly, they mainly corresponded to the 3' untranslated region of the mRNA. In part, the prevalence of EST sequences derived from the 3' end of the mRNA is a result of the fact that typical techniques for obtaining cDNAs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs (Adams *et al.*, *Nature* 377:3-174, 1996, Hillier *et al.*, *Genome Res.* 6:807-828, 1996). In addition, in those reported instances where longer cDNA sequences have been obtained, the reported sequences typically correspond to coding sequences and do not include the full 5' untranslated region (5'UTR) of the mRNA from which the cDNA is derived. Indeed, 5'UTRs have been shown to affect either the stability or translation of mRNAs. Thus, regulation of gene expression may be achieved through the use of alternative 5'UTRs as shown, for instance, for the translation of the tissue inhibitor of metalloprotease mRNA in mitogenically activated cells (Waterhouse *et al.*, *J Biol Chem.* 265:5585-9, 1990). Furthermore, modification of 5'UTR through mutation, insertion or translocation events may even be implied in pathogenesis. For instance, the Fragile X syndrome, the most common cause of inherited mental retardation, is partly due to an insertion of multiple CGG trinucleotides in the 5'UTR of the Fragile X mRNA resulting in the inhibition of protein synthesis via ribosome stalling (Feng *et al.*, *Science* 268:731-4, 1995). An aberrant mutation in regions of the 5'UTR known to inhibit translation of the proto-oncogene *c-myc* was shown to result in upregulation of c-myc protein levels in cells derived from patients with multiple myelomas (Willis *et al.*, *Curr Top Microbiol Immunol* 224:269-76, 1997). In addition, the

use of oligo-dT primed cDNA libraries does not allow the isolation of complete 5'UTRs since such incomplete sequences obtained by this process may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain
5 sequences derived from the 5' ends of mRNAs.

Moreover, despite the great amount of EST data that large-scale sequencing projects have yielded (Adams *et al.*, *Nature* 377:174, 1996; Hillier *et al.*, *Genome Res.* 6:807-828, 1996), information concerning the biological function of the mRNAs corresponding to such obtained cDNAs has revealed to be limited. Indeed, whereas the knowledge of the complete coding
10 sequence is absolutely necessary to investigate the biological function of mRNAs, ESTs yield only partial coding sequences. So far, large-scale full-length cDNA cloning has been achieved only with limited success because of the poor efficiency of methods for constructing full-length cDNA libraries. Indeed, such methods require either a large amount of mRNA (Ederly *et al.*, 1995), thus
15 resulting in non representative full-length libraries when small amounts of tissue are available or require PCR amplification (Maruyama *et al.*, 1994; CLONTECHniques, 1996) to obtain a reasonable number of clones, thus yielding strongly biased cDNA libraries where rare and long cDNAs are lost. Thus, there is a need to obtain full-length cDNAs, *i.e.* cDNAs containing the full coding sequence of their corresponding mRNAs.

While many sequences derived from human chromosomes have practical applications,
20 approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. Of the 30,000-120,000 protein coding genes, those genes encoding proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often involved in cell to cell communication and
25 may be responsible for producing a clinically relevant response in their target cells. In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon- α , interferon- β , interferon- γ , and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney
30 carcinoma, chemotherapy induced neutropenia and multiple sclerosis. For these reasons, cDNAs encoding secreted proteins or fragments thereof represent a particularly valuable source of therapeutic agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

In addition to being therapeutically useful themselves, secretory proteins include short
35 peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are encoded by the signal sequences located at the 5' ends of the coding sequences of genes encoding secreted proteins. Because these signal peptides will direct the extracellular secretion of any protein to which they are operably linked, the signal sequences may be exploited to direct the efficient secretion of any protein by operably linking the signal sequences to a gene encoding the

protein for which secretion is desired. In addition, fragments of the signal peptides called membrane-translocating sequences may also be used to direct the intracellular import of a peptide or protein of interest. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cells in which it is produced. Signal sequences encoding signal peptides also find application in simplifying protein purification techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired protein must be selected. Thus, there exists a need to identify and characterize the 5' fragments of the genes for secretory proteins which encode signal peptides.

Sequences coding for secreted proteins may also find application as therapeutics or diagnostics. In particular, such sequences may be used to determine whether an individual is likely to express a detectable phenotype, such as a disease, as a consequence of a mutation in the coding sequence for a secreted protein. In instances where the individual is at risk of suffering from a disease or other undesirable phenotype as a result of a mutation in such a coding sequence, the undesirable phenotype may be corrected by introducing a normal coding sequence using gene therapy. Alternatively, if the undesirable phenotype results from overexpression of the protein encoded by the coding sequence, expression of the protein may be reduced using antisense or triple helix based strategies.

The secreted human polypeptides encoded by the coding sequences may also be used as therapeutics by administering them directly to an individual having a condition, such as a disease, resulting from a mutation in the sequence encoding the polypeptide. In such an instance, the condition can be cured or ameliorated by administering the polypeptide to the individual.

In addition, the secreted human polypeptides or fragments thereof may be used to generate antibodies useful in determining the tissue type or species of origin of a biological sample. The antibodies may also be used to determine the cellular localization of the secreted human polypeptides or the cellular localization of polypeptides which have been fused to the human polypeptides. In addition, the antibodies may also be used in immunoaffinity chromatography techniques to isolate, purify, or enrich the human polypeptide or a target polypeptide which has been fused to the human polypeptide.

Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross *et al.*, *Nature Genetics* 6: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein (Mortlock *et al.*, *Genome Res.* 6:327-335, 1996). Both of these approaches have their limits due to a lack of specificity and of

comprehensiveness. Thus, there exists a need to identify and systematically characterize the 5' fragments of the genes.

cDNAs including the 5' ends of their corresponding mRNA may be used to efficiently identify and isolate 5'UTRs and upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA (Theil *et al.*, *BioFactors* 4:87-93, (1993). Once identified and characterized, these regulatory regions may be utilized in gene therapy or protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, cDNAs containing the 5' ends of secretory protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5' coding sequences of genes encoding secretory proteins.

SUMMARY OF THE INVENTION

The present invention provides a purified or isolated polynucleotide comprising, consisting of, or consisting essentially of a nucleotide sequence selected from the group consisting of: (a) the sequences of SEQ ID NOs:1-169, 339-455, 561-784; (b) the sequences of clone inserts of the deposited clone pool; (c) the coding sequences of SEQ ID NOs:1-169, 339-455, 561-784; (d) the coding sequences of the clone inserts of the deposited clone pool; (e) the sequences encoding one of the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918; (f) the sequences encoding one of the polypeptides encoded by the clone inserts of the deposited clone pool; (g) the genomic sequences coding for the GENSET polypeptides; (h) the 5' transcriptional regulatory regions of GENSET genes; (i) the 3' transcriptional regulatory regions of GENSET genes; (j) the polynucleotides comprising the nucleotide sequence of any combination of (g)-(i); (k) the variant polynucleotides of any of the polynucleotides of (a)-(j); (l) the polynucleotides comprising a nucleotide sequence of (a)-(k), wherein the polynucleotide is single stranded, double stranded, or a portion is single stranded and a portion is double stranded; (m) the polynucleotides comprising a nucleotide sequence complementary to any of the single stranded polynucleotides of (l). The invention further provides for fragments of the nucleic acid molecules of (a)-(m) described above.

Further embodiments of the invention include purified or isolated polynucleotides that comprise, consist of, or consist essentially of a nucleotide sequence at least 70% identical, more preferably at least 75%, and even more preferably at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical, to any of the nucleotide sequences in (a)-(m) above, e.g. over a region of at least about 25, 50, 100, 150, 250, 500, 1000, or more contiguous nucleotides, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a)-(m) above.

The present invention also relates to recombinant vectors, which include the purified or isolated polynucleotides of the present invention, and to host cells recombinant for the polynucleotides of the present invention, as well as to methods of making such vectors and host

cells. The present invention further relates to the use of these recombinant vectors and recombinant host cells in the production of GENSET polypeptides.

The invention further provides a purified or isolated polypeptide comprising, consisting of, or consisting essentially of an amino acid sequence selected from the group consisting of: (a) the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918; (b) the polypeptides encoded by the clone inserts of the deposited clone pool; (c) the epitope-bearing fragments of the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918; (d) the epitope-bearing fragments of the polypeptides encoded by the clone inserts contained in the deposited clone pool; (e) the domains of the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918; (f) the domains of the polypeptides encoded by the clone inserts contained in the deposited clone pool; and (g) the allelic variant polypeptides of any of the polypeptides of (a)-(f). The invention further provides for fragments of the polypeptides of (a)-(g) above, such as those having biological activity or comprising biologically functional domain(s).

The present invention further includes polypeptides with an amino acid sequence with at least 70% similarity, and more preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% similarity to those polypeptides described in (a)-(g), as well as polypeptides having an amino acid sequence at least 70% identical, more preferably at least 75% identical, and still more preferably 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to those polypeptides described in (a)-(g), e.g. over a region of at least about 25, 50, 100, 150, 250, 500, 1000, or more amino acids.

The invention further relates to methods of making the polypeptides of the present invention.

The present invention further relates to transgenic plants or animals, wherein said transgenic plant or animal is transgenic for a polynucleotide of the present invention and expresses a polypeptide of the present invention.

The invention further relates to antibodies that specifically bind to GENSET polypeptides of the present invention and fragments thereof as well as to methods for producing such antibodies and fragments thereof.

The invention also provides kits, uses and methods for detecting GENSET gene expression and/or biological activity in a biological sample. One such method involves assaying for the expression of a GENSET polynucleotide in a biological sample using the polymerase chain reaction (PCR) to amplify and detect GENSET polynucleotides or Southern and Northern blot hybridization to detect GENSET genomic DNA, cDNA or mRNA. Alternatively, a method of detecting GENSET gene expression in a test sample can be accomplished using a compound which binds to a GENSET polypeptide of the present invention or a portion of a GENSET polypeptide.

The present invention also relates to diagnostic methods and uses of GENSET polynucleotides and polypeptides for identifying individuals or non-human animals having elevated or reduced levels of GENSET gene products, which individuals are likely to benefit from therapies to suppress or enhance GENSET gene expression, respectively, and to methods of identifying individuals or non-human animals at increased risk for developing, or at present having, certain diseases/disorders associated with GENSET polypeptide expression or biological activity.

The present invention also relates to kits, uses and methods of screening compounds for their ability to modulate (e.g. increase or inhibit) the activity or expression of GENSET polypeptides including compounds that interact with GENSET gene regulatory sequences and compounds that interact directly or indirectly with a GENSET polypeptide. Uses of such
5 compounds are also within the scope of the present invention.

The present invention also relates to pharmaceutical or physiologically acceptable compositions comprising, an active agent, the polypeptides, polynucleotides or antibodies of the present invention, as well as, typically, a pharmaceutically acceptable carrier.

The present invention also relates to computer systems containing cDNA codes and
10 polypeptide codes of sequences of the invention and to computer-related methods of comparing sequences, identifying homology or features using GENSET polypeptides or GENSET polynucleotide sequences of the invention.

In another aspect, the present invention provides an isolated polynucleotide, the polynucleotide comprising a nucleic acid sequence encoding: i) a polypeptide comprising an amino
15 acid sequence having at least about 80% identity to any one of the sequences shown as SEQ ID NOs:170-338, 456-560, 785-918 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool; or a biologically active fragment of the polypeptide.

In one embodiment, the polypeptide comprises any one of the sequences shown as SEQ ID NOs:170-338, 456-560, 785-918 or any one of the sequences of the polypeptides encoded by the
20 clone inserts of the deposited clone pool. In another embodiment, the polypeptide comprises a signal peptide. In another embodiment, the polypeptide is a mature protein. In another embodiment, the nucleic acid sequence has at least about 80% identity over at least about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-169, 339-455, 561-784 or any one of the sequences of the clone inserts of the deposited clone pool. In another
25 embodiment, the polynucleotide hybridizes under stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID NOs:1-169, 339-455, 561-784 or any one of the sequences of the clone inserts of the deposited clone pool. In another embodiment, the nucleic acid sequence comprises any one of the sequences shown as SEQ ID NOs:1-169, 339-455, 561-784 or any one the sequences of the clone inserts of the deposited clone pool. In another embodiment,
30 the polynucleotide is operably linked to a promoter.

In another aspect, the present invention provides an expression vector comprising any of the herein-described polynucleotides, operably linked to a promoter. In another aspect, the present invention provides a host cell recombinant for any of the herein-described polynucleotides. In another aspect, the present invention provides a non-human transgenic animal comprising the host
35 cell.

In another aspect, the present invention provides a method of making a GENSET polypeptide, the method comprising a) providing a population of host cells comprising a herein-described polynucleotide and b) culturing the population of host cells under conditions conducive to the production of the polypeptide within said host cells.

In one embodiment, the method further comprises purifying the polypeptide from the population of host cells.

In another aspect, the present invention provides a method of making a GENSET polypeptide, the method comprising a) providing a population of cells comprising a polynucleotide
5 encoding a herein-described polypeptide; b) culturing the population of cells under conditions conducive to the production of the polypeptide within the cells; and c) purifying the polypeptide from the population of cells.

In another aspect, the present invention provides an isolated polynucleotide, the polynucleotide comprising a nucleic acid sequence having at least about 80% identity over at least
10 about 100 contiguous nucleotides to any one of the sequences shown as SEQ ID NOs:1-169, 339-455, 561-784 or any one of the sequences of the clone inserts of the deposited clone pool.

In one embodiment, the polynucleotide hybridizes under stringent conditions to a polynucleotide comprising any one of the sequences shown as SEQ ID NOs:1-169, 339-455, 561-784 or any one of the sequences of the clone inserts of the deposited clone pool. In another
15 embodiment, the polynucleotide comprises any one of the sequences shown as SEQ ID NOs:1-169, 339-455, 561-784 or any one of the sequences of the clone inserts of the deposited clone pool.

In another aspect, the present invention provides a biologically active polypeptide encoded by any of the herein-described polynucleotides.

In another aspect, the present invention provides an isolated polypeptide or biologically
20 active fragment thereof, the polypeptide comprising an amino acid sequence having at least about 80% sequence identity to any one of the sequences shown as SEQ ID NOs:170-338, 456-560, 785-918 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool.

In one embodiment, the polypeptide is selectively recognized by an antibody raised against
25 an antigenic polypeptide, or an antigenic fragment thereof, the antigenic polypeptide comprising any one of the sequences shown as SEQ ID NOs:170-338, 456-560, 785-918 or any one of the sequences of polypeptides encoded by the clone inserts of the deposited clone pool. In another embodiment, the polypeptide comprises any one of the sequences shown as SEQ ID NOs:170-338, 456-560, 785-918 or any one of the sequences of polypeptides encoded by the clone inserts of the
30 deposited clone pool. In another embodiment, the polypeptide comprises a signal peptide. In another embodiment, the polypeptide is a mature protein.

In another aspect, the present invention provides an antibody that specifically binds to any of the herein-described polypeptides.

In another aspect, the present invention provides a method of determining whether a
35 GENSET gene is expressed within a mammal, the method comprising the steps of: a) providing a biological sample from said mammal; b) contacting said biological sample with either of: i) a polynucleotide that hybridizes under stringent conditions to any of the herein-described polynucleotides; or ii) a polypeptide that specifically binds to any of the herein-described polypeptides; and c) detecting the presence or absence of hybridization between the polynucleotide

and an RNA species within the sample, or the presence or absence of binding of the polypeptide to a protein within the sample; wherein a detection of the hybridization or of the binding indicates that the GENSET gene is expressed within the mammal.

In one embodiment, the polynucleotide is a primer, and the hybridization is detected by
5 detecting the presence of an amplification product comprising the sequence of the primer. In another embodiment, the polypeptide is an antibody.

In another aspect, the present invention provides a method of determining whether a mammal has an elevated or reduced level of GENSET gene expression, the method comprising the steps of: a) providing a biological sample from the mammal; and b) comparing the amount of any of
10 the herein-described polypeptides, or of an RNA species encoding the polypeptide, within the biological sample with a level detected in or expected from a control sample; wherein an increased amount of the polypeptide or the RNA species within the biological sample compared to the level detected in or expected from the control sample indicates that the mammal has an elevated level of the GENSET gene expression, and wherein a decreased amount of the polypeptide or the RNA
15 species within the biological sample compared to the level detected in or expected from the control sample indicates that the mammal has a reduced level of the GENSET gene expression.

In another aspect, the present invention provides a method of identifying a candidate modulator of a GENSET polypeptide, the method comprising: a) contacting any of the herein-described polypeptides with a test compound; and b) determining whether the compound
20 specifically binds to the polypeptide; wherein a detection that the compound specifically binds to the polypeptide indicates that the compound is a candidate modulator of the GENSET polypeptide.

In one embodiment, the method further comprises testing the biological activity of the GENSET polypeptide in the presence of the candidate modulator, wherein an alteration in the biological activity of the GENSET polypeptide in the presence of the compound in comparison to
25 the activity in the absence of said compound indicates that the compound is a modulator of the GENSET polypeptide.

In another aspect, the present invention provides a method for the production of a pharmaceutical composition, the method comprising a) identifying a modulator of a GENSET polypeptide using any of the herein-described methods; and b) combining the modulator with a
30 pharmaceutically acceptable carrier.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is a block diagram of an exemplary computer system.

Figure 2 is a flow diagram illustrating one embodiment of a process 200 for comparing a
35 new nucleotide or protein sequence with a database of sequences in order to determine the identity levels between the new sequence and the sequences in the database.

Figure 3 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous.

Figure 4 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

BRIEF DESCRIPTION OF THE TABLES

5 Table I provides the SEQ ID Nos in the present application (with the SEQ ID Nos corresponding to nucleic acid sequences preceded by "NUC", and the SEQ ID Nos corresponding to the encoded polypeptide sequences preceded by "PRT") that correspond to a SEQ ID NO in priority application number 60/197,873. Applicants' internal designation number (Clone ID) corresponding to each sequence identification (SEQ ID) number is also provided.

10 Table II lists the putative chromosomal location of the polynucleotides of the present invention. The SEQ ID NO listed for each polynucleotide is that from the priority application 60/197,873; the corresponding SEQ ID NOs for the sequence in the present application can be determined by referring to Table I.

15 Table III lists the number of hits in Genset's cDNA libraries of tissues and cell types for polynucleotides of the invention. The following abbreviations are used to refer to each cell or tissue type: A=Brain; B=Fetal brain; C=Fetal kidney; D=Fetal liver; E=Pituitary gland; F=Liver; G=Placenta; H=Prostate; I=Salivary gland; J=Stomach/Intestine; and K=Testis. The SEQ ID NO listed for each polynucleotide is that from the priority application 60/197,873; the corresponding SEQ ID NOs for the sequence in the present application can be determined by referring to Table I.

20 Table IV lists the number of hits in publicly available library of tissues and cell types for polynucleotides of the invention. The SEQ ID NO listed for each polynucleotide is that from the priority application 60/197,873; the corresponding SEQ ID NOs for each sequence in the present application can be determined by referring to Table I.

25 Table V lists the tissues and cell types in which the polynucleotide sequences of the present invention are over- or under-represented. The SEQ ID NO listed for each polynucleotide is that from the priority application 60/197,873; the corresponding SEQ ID NOs for each sequence in the present application can be determined by referring to Table I.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

30 SEQ ID NOs:1-169, 339-455, 561-784 are the nucleotide sequences of cDNAs, with open reading frames as indicated as features (CDS). When appropriate, the locations of the potential polyadenylation site and polyadenylation signal are also indicated.

 SEQ ID NOs:170-338, 456-560, 785-918 are the amino acid sequences of proteins encoded by the cDNAs of SEQ ID NOs:1-169, 339-455, 561-784.

35 SEQ ID NOs:1-85, 339-400, 406-407, 413-415, 561-594, and 634-651 are the nucleotide sequences of cDNAs encoding a potentially secreted protein. The locations of the ORFs and sequences encoding signal peptides are listed in the accompanying Sequence Listing. In addition,

the von Heijne score of the signal peptide computed as described below is listed as the "score" in the accompanying Sequence Listing. The sequence of the signal-peptide is listed as "seq" in the accompanying Sequence Listing. The "/" in the signal peptide sequence indicates the location where proteolytic cleavage of the signal peptide occurs to generate a mature protein. When
 5 appropriate, the locations of the first and last nucleotides of the coding sequences, eventually the locations of the first and last nucleotides of the polyA and the locations of the first and last nucleotides of the polyA sites are indicated.

SEQ ID NOs:86-169, 401-405, 408-412, 416-455, 595-633, 652-784 are the nucleotide sequences of cDNAs in which no sequence encoding a signal peptide has been identified to date.

10 However, it remains possible that subsequent analysis will identify a sequence encoding a signal peptide in these nucleic acids. The locations of the ORFs are listed in the accompanying Sequence Listing. When appropriate, the locations of the first and last nucleotides of the coding sequences, eventually the locations of the first and last nucleotides of the polyA and the locations of the first and last nucleotides of the polyA sites are indicated.

15 SEQ ID NOs:170-254, 456-517, 520-521, 527-529, 785-818, and 858-875 are the amino acid sequences of polypeptides which contain a signal peptide. These polypeptides are encoded by the cDNAs of SEQ ID NOs: 1-85, 339-400, 406-407, 413-415, 561-594, and 634-651. The location of the signal peptide is listed in the accompanying Sequence Listing.

20 SEQ ID NOs:255-338, 517-519, 522-526, 530-560, 819-857, 876-918 are the amino acid sequences of polypeptides in which no signal peptide has been identified to date. However, it remains possible that subsequent analysis will identify a signal peptide in these polypeptides. These polypeptides are encoded by the nucleic acids of SEQ ID NOs: 86-169, 401-405, 408-412, 416-455, 595-633, 652-784.

In accordance with the regulations relating to Sequence Listings, the following codes have
 25 been used in the Sequence Listing to describes nucleotide sequences. The code "r" in the sequences indicates that the nucleotide may be a guanine or an adenine. The code "y" in the sequences indicates that the nucleotide may be a thymine or a cytosine. The code "m" in the sequences indicates that the nucleotide may be an adenine or a cytosine. The code "k" in the sequences indicates that the nucleotide may be a guanine or a thymine. The code "s" in the sequences
 30 indicates that the nucleotide may be a guanine or a cytosine. The code "w" in the sequences indicates that the nucleotide may be an adenine or an thymine. In addition, all instances of the symbol "n" in the nucleic acid sequences mean that the nucleotide can be adenine, guanine, cytosine or thymine.

In some instances, the polypeptide sequences in the Sequence Listing contain the symbol
 35 "Xaa." These "Xaa" symbols indicate either (1) a residue which cannot be identified because of nucleotide sequence ambiguity or (2) a stop codon in the determined sequence where applicants believe one should not exist (if the sequence were determined more accurately). In some instances, several possible identities of the unknown amino acids may be suggested by the genetic code.

In the case of secreted proteins, it should be noted that, in accordance with the regulations governing Sequence Listings, in the appended Sequence Listing the encoded protein (i.e. the protein containing the signal peptide and the mature protein or part thereof) extends from an amino acid residue having a negative number through a positively numbered amino acid residue. Thus, the first amino acid of the mature protein resulting from cleavage of the signal peptide is designated as amino acid number 1, and the first amino acid of the signal peptide is designated with the appropriate negative number.

DETAILED DESCRIPTION OF THE INVENTION AND PREFERRED EMBODIMENTS

DEFINITIONS

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

The term "GENSET gene," when used herein, encompasses genomic, mRNA and cDNA sequences encoding a GENSET polypeptide, including the 5' and 3' untranslated regions of said sequences.

The term "GENSET polypeptide biological activity" or "GENSET biological activity" is intended for polypeptides exhibiting any activity similar, but not necessarily identical, to an activity of a GENSET polypeptide of the invention. The GENSET polypeptide biological activity of a given polypeptide may be assessed using any suitable biological assay, a number of which are known to those skilled in the art. In contrast, the term "biological activity" refers to any activity that any polypeptide may have.

The term "corresponding mRNA" refers to mRNA which was or can be a template for cDNA synthesis for producing a cDNA of the present invention.

The term "corresponding genomic DNA" refers to genomic DNA which encodes an mRNA of interest, e.g. corresponding to a cDNA of the invention, which genomic DNA includes the sequence of one of the strands of the mRNA, in which thymidine residues in the sequence of the genomic DNA (or cDNA) are replaced by uracil residues in the mRNA.

The term "deposited clone pool" is used herein to refer to the pool of clones entitled cDNA-8-2000, deposited with the ATCC on September 27, 2000, or the pool of clones entitled cDNA-11-2000, deposited with the ATCC on November 27, 2000, or any other deposited clone pool containing a clone corresponding to any of the herein-described sequences.

The term "heterologous", when used herein, is intended to designate any polynucleotide or polypeptide other than a GENSET polynucleotide or GENSET polypeptide of the invention, respectively.

"Providing" with respect to, e.g. a biological sample, population of cells, etc. indicates that the sample, population of cells, etc. is somehow used in a method or procedure. Significantly,

“providing” a biological sample or population of cells does not require that the sample or cells are specifically isolated or obtained for the purposes of the invention, but can instead refer, for example, to the use of a biological sample obtained by another individual, for another purpose.

An “amplification product” refers to a product of any amplification reaction, e.g. PCR, RT-PCR, LCR, etc.

A “modulator” of a protein or other compound refers to any agent that has a functional effect on the protein, including physical binding to the protein, alterations of the quantity or quality of expression of the protein, altering any measurable or detectable activity, property, or behavior of the protein, or in any way interacts with the protein or compound.

“A test compound” can be any molecule that is evaluated for its ability to modulate a protein or other compound.

An antibody or other compound that specifically binds to a polypeptide or polynucleotide of the invention is also said to “selectively recognize” the polypeptide or polynucleotide.

The term “isolated” with respect to a molecule requires that the molecule be removed from its original environment (e. g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment. For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated. Specifically excluded from the definition of “isolated” are: naturally-occurring chromosomes (such as chromosome spreads), artificial chromosome libraries, genomic libraries, and cDNA libraries that exist either as an *in vitro* nucleic acid preparation or as a transfected/transformed host cell preparation, wherein the host cells are either an *in vitro* heterogeneous preparation or plated as a heterogeneous population of single colonies. Also specifically excluded are the above libraries wherein a specified polynucleotide makes up less than 5% of the number of nucleic acid inserts in the vector molecules. Further specifically excluded are whole cell genomic DNA or whole cell RNA preparations (including said whole cell preparations which are mechanically sheared or enzymatically digested). Further specifically excluded are the above whole cell preparations as either an *in vitro* preparation or as a heterogeneous mixture separated by electrophoresis (including blot transfers of the same) wherein the polynucleotide of the invention has not further been separated from the heterologous polynucleotides in the electrophoresis medium (e.g., further separating by excising a single band from a heterogeneous band population in an agarose gel or nylon blot).

The term “purified” does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. As an example, purification from 0.1% concentration to 10% concentration is two

orders of magnitude. To illustrate, individual cDNA clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^4 - 10^6 fold purification of the native message.

10 The term "purified" is further used herein to describe a polypeptide or polynucleotide of the invention which has been separated from other compounds including, but not limited to, polypeptides or polynucleotides, carbohydrates, lipids, etc. The term "purified" may be used to specify the separation of monomeric polypeptides of the invention from oligomeric forms such as homo- or hetero- dimers, trimers, etc. The term "purified" may also be used to specify the separation of covalently closed (i.e. circular) polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polypeptide or polynucleotide typically comprises about 50%, preferably 60 to 90% weight/weight of a polypeptide or polynucleotide sample, respectively, more usually about 95%, and preferably is over about 99% pure. Polypeptide and polynucleotide purity, or homogeneity, is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art. As an alternative embodiment, purification of the polypeptides and polynucleotides of the present invention may be expressed as "at least" a percent purity relative to heterologous polypeptides and polynucleotides (DNA, RNA or both). As a preferred embodiment, the polypeptides and polynucleotides of the present invention are at least; 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 96%, 98%, 99%, or 100% pure relative to heterologous polypeptides and polynucleotides, respectively. As a further preferred embodiment the polypeptides and polynucleotides have a purity ranging from any number, to the thousandth position, between 90% and 100% (e.g., a polypeptide or polynucleotide at least 99.995% pure) relative to either heterologous polypeptides or polynucleotides, respectively, or as a weight/weight ratio relative to all compounds and molecules other than those existing in the carrier. Each number representing a percent purity, to the thousandth position, may be claimed as individual species of purity.

35 As used interchangeably herein, the terms "nucleic acid molecule(s)", "oligonucleotide(s)", and "polynucleotide(s)" include RNA or DNA (either single or double stranded, coding, complementary or antisense), or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form (although each of the above species may be particularly specified). The term "nucleotide" is used herein as an adjective to describe molecules comprising RNA, DNA, or

RNA/DNA hybrid sequences of any length in single-stranded or duplex form. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. The term "nucleotide" is also used

5 herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. The term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modification such as (a) an

10 alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar. For examples of analogous linking groups, purine, pyrimidines, and sugars, see, for example, PCT publication No. WO 95/04064, which disclosure is hereby incorporated by reference in its entirety. Preferred modifications of the present invention include, but are not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-

15 acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-

20 methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v) ybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, and 2,6-diaminopurine. The polynucleotide sequences of the invention may be prepared by any known method, including

25 synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art. Methylenemethylimino linked oligonucleosides as well as mixed backbone compounds having, may be prepared as described in U.S. Pat. Nos. 5,378,825; 5,386,023; 5,489,677; 5,602,240; and 5,610,289, which disclosures are hereby incorporated by reference in their entirety. Formacetal and thioformacetal linked oligonucleosides may be prepared

30 as described in U.S. Pat. Nos. 5,264,562 and 5,264,564, which disclosures are hereby incorporated by reference in their entirety. Ethylene oxide linked oligonucleosides may be prepared as described in U.S. Pat. No. 5,223,618, which disclosure is hereby incorporated by reference in its entirety. Phosphinate oligonucleotides may be prepared as described in U.S. Pat. No. 5,508,270, which disclosure is hereby incorporated by reference in its entirety. Alkyl phosphonate

35 oligonucleotides may be prepared as described in U.S. Pat. No. 4,469,863, which disclosure is hereby incorporated by reference in its entirety. 3'-Deoxy-3'-methylene phosphonate oligonucleotides may be prepared as described in U.S. Pat. Nos. 5,610,289 or 5,625,050 which disclosures are hereby incorporated by reference in their entirety. Phosphoramidite oligonucleotides may be prepared as described in U.S. Pat. No. 5,256,775 or U.S. Pat. No.

5,366,878 which disclosures are hereby incorporated by reference in their entireties.

Alkylphosphonothioate oligonucleotides may be prepared as described in published PCT applications WO 94/17093 and WO 94/02499 which disclosures are hereby incorporated by reference in their entireties. 3'-Deoxy-3'-amino phosphoramidate oligonucleotides may be prepared as described in U.S. Pat. No. 5,476,925, which disclosure is hereby incorporated by reference in its entirety. Phosphotriester oligonucleotides may be prepared as described in U.S. Pat. No. 5,023,243, which disclosure is hereby incorporated by reference in its entirety. Borano phosphate oligonucleotides may be prepared as described in U.S. Pat. Nos. 5,130,302 and 5,177,198 which disclosures are hereby incorporated by reference in their entireties.

10 The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another by virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (see Stryer, 1995, which disclosure is hereby incorporated by reference in its entirety).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G. "Complement" is used herein as a synonym from "complementary polynucleotide", "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind. Unless otherwise stated, all complementary polynucleotides are fully complementary on the whole length of the considered polynucleotide.

The terms "polypeptide" and "protein", used interchangeably herein, refer to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude chemical or post-expression modifications of the polypeptides of the invention, although chemical or post-expression modifications of these polypeptides may be included excluded as specific embodiments. Therefore, for example, modifications to polypeptides that include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Further, polypeptides with these modifications may be specified as individual species to be included or excluded from the present invention. The natural or other chemical modifications, such as those listed in examples above can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or

carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Modifications include

5 acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI
10 anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance Creighton (1993); Seifter *et al.*, (1990); Rattan *et al.*, (1992)). Also included within the definition are polypeptides which contain one or more analogs of an amino acid
15 (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems, etc...), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

As used herein, the terms "recombinant polynucleotide" and "polynucleotide construct" are
20 used interchangeably to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment. In particular, these terms mean that the polynucleotide or cDNA is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the cDNAs will represent 5% or
25 more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched cDNAs represent 15% or more of the number of nucleic acid inserts in the population of
30 recombinant backbone molecules. More preferably, the enriched cDNAs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched cDNAs represent 90% or more (including any number between 90 and 100%, to the thousandth position, e.g., 99.5%) of the number of nucleic acid inserts in the population of recombinant backbone molecules.

35 The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

As used herein, the term “operably linked” refers to a linkage of polynucleotide elements in a functional relationship. A sequence which is “operably linked” to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest.

- 5 For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence.

- As used herein, the term “non-human animal” refers to any non-human animal, including insects, birds, rodents and more usually mammals. Preferred non-human animals include: primates; farm animals such as swine, goats, sheep, donkeys, cattle, horses, chickens, rabbits; and rodents,
 10 preferably rats or mice. As used herein, the term “animal” is used to refer to any species in the animal kingdom, preferably vertebrates, including birds and fish, and more preferable a mammal. Both the terms “animal” and “mammal” expressly embrace human subjects unless preceded with the term “non-human”.

- The term “domain” refers to an amino acid fragment with specific biological properties.
 15 This term encompasses all known structural and linear biological motifs. Examples of such motifs include but are not limited to leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal peptides which direct the secretion of proteins, sites for post-translational modification, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

- 20 Although each of these terms has a distinct meaning, the terms "comprising", "consisting of" and "consisting essentially of" may be interchanged for one another throughout the instant application. The term "having" has the same meaning as "comprising" and may be replaced with either the term "consisting of" or "consisting essentially of".

- Unless otherwise specified in the application, nucleotides and amino acids of
 25 polynucleotides and polypeptides, respectively, of the present invention are contiguous and not interrupted by heterologous sequences.

Identity Between Nucleic Acids Or Polypeptides

- The terms “percentage of sequence identity” and “percentage homology” are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are
 30 determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue
 35 occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs

include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, CLUSTALW, FASTDB (Pearson and Lipman, 1988; Altschul *et al.*, 1990; Thompson *et al.*, 1994; Higgins *et al.*, 1996; Altschul *et al.*, 1990; Altschul *et al.*, 1993; Brutlag *et al.*, 1990), the disclosures of which are incorporated by reference in their entirety.

5 In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul *et al.*, 1990, 1993, 1997), the disclosures of which are incorporated by reference in their entirety. In particular, five specific BLAST programs are used to perform the following task:

10 (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide
15 sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

20 The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62
25 matrix (Gonnet *et al.*, 1992; Henikoff and Henikoff, 1993, the disclosures of which are incorporated by reference in their entirety). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, the disclosure of which is incorporated by reference in its entirety). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-
30 specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990), the disclosure of which is incorporated by reference in its entirety. The BLAST programs may be used with the default parameters or with modified parameters provided by the user.

35 Another preferred method for determining the best overall match between a query nucleotide sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag *et al.* (1990), the disclosure of which is incorporated by reference in its entirety. In a sequence alignment the query and subject sequences are both DNA sequences. An

RNA sequence can be compared by first converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix = Unitary, k-tuple = 4, Mismatch Penalty = 1, Joining Penalty = 30, Randomization Group Length = 0, Cutoff Score = 1, Gap Penalty = 5, Gap Size Penalty = 0.05, Window Size = 500 or the length of the subject nucleotide sequence, whichever is shorter. If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using 10, the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only nucleotides outside the 5' and 3' nucleotides of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score. For example, a 90 nucleotide subject sequence is aligned to a 100 nucleotide query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 nucleotides at 5' end. The 10 unpaired nucleotides represent 10% of the sequence (number of nucleotides at the 5' and 3' ends not matched/total number of nucleotides in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 nucleotides were perfectly matched the final percent identity would be 90%. In another example, a 90 nucleotide subject sequence is compared with a 100 nucleotide query sequence. This time the deletions are internal deletions so that there are no nucleotides on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only nucleotides 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected. No other manual corrections are made for the purposes of the present invention.

Another preferred method for determining the best overall match between a query amino acid sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag *et al.* (1990). In a sequence alignment the query and subject sequences are both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix = PAM 0, k-tuple = 2, Mismatch Penalty = 1, Joining Penalty = 20, Randomization Group Length = 0, Cutoff Score = 1, Window Size = sequence length, Gap Penalty = 5, Gap Size Penalty = 0.05, Window Size = 500 or

the length of the subject amino acid sequence, whichever is shorter. If the subject sequence is shorter than the query sequence due to N-or C-terminal deletions, not because of internal deletions, the results, in percent identity, must be manually corrected. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C- terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query amino acid residues outside the farthest N- and C-terminal residues of the subject sequence. For example, a 90 amino acid residue subject sequence is aligned with a 100-residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not match/align with the first residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C-termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90-residue subject sequence is compared with a 100-residue query sequence. This time the deletions are internal so there are no residues at the N- or C-termini of the subject sequence, which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected. No other manual corrections are made for the purposes of the present invention.

The term "percentage of sequence similarity" refers to comparisons between polypeptide sequences and is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which an identical or equivalent amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence similarity. Similarity is evaluated using any of the variety of sequence comparison algorithms and programs known in the art, including those

described above in this section. Equivalent amino acid residues are defined herein in the "Mutated polypeptides" section.

POLYNUCLEOTIDES OF THE INVENTION

5 The present invention concerns GENSET genomic and cDNA sequences. The present invention encompasses GENSET genes, polynucleotides comprising GENSET genomic and cDNA sequences, as well as fragments and variants thereof. These polynucleotides may be purified, isolated, or recombinant.

Also encompassed by the present invention are allelic variants, orthologs, splice variants,
10 and/or species homologues of the GENSET genes. Procedures known in the art can be used to obtain full-length genes and cDNAs, allelic variants, splice variants, full-length coding portions, orthologs, and/or species homologues of genes and cDNAs corresponding to a nucleotide sequence selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool, using information from the sequences
15 disclosed herein or the clone pool deposited with the ATCC or other depositary authority. For example, allelic variants, orthologs and/or species homologues may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants and/or the desired homologue using any technique known to those skilled in the art including those described into the section entitled "To find similar
20 sequences".

In a specific embodiment, the polynucleotides of the invention are at least 15, 30, 50, 100, 125, 500, or 1000 continuous nucleotides. In another embodiment, the polynucleotides are less than or equal to 300kb, 200kb, 100kb, 50kb, 10kb, 7.5kb, 5kb, 2.5kb, 2kb, 1.5kb, or 1kb in length. In a further embodiment, polynucleotides of the invention comprise a portion of the coding sequences,
25 as disclosed herein, but do not comprise all or a portion of any intron. In another embodiment, the polynucleotides comprising coding sequences do not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene of interest in the genome). In other embodiments, the polynucleotides of the invention do not contain the coding sequence of more than 1000, 500, 250, 100, 75, 50, 25, 20, 15, 10, 5, 4, 3, 2, or 1 naturally occurring genomic flanking gene(s).

30 Deposited clone pool of the invention

Expression of GENSET genes has been shown to lead to the production of at least one mRNA species per GENSET gene, which cDNA sequence is set forth in the appended sequence listing as SEQ ID NOs:1-169, 339-455, 561-784. The cDNAs (SEQ ID NOs:1-169, 339-455, 561-784) corresponding to these GENSET mRNA species were cloned either in the vector pBluescriptII
35 SK⁻ (Stratagene) or in a vector called pPT. Cells containing the cloned cDNAs of the present invention are maintained in permanent deposit by the inventors at Genset, S.A., 24 Rue Royale, 75008 Paris, France. Each cDNA can be removed from the Bluescript vector in which it was

inserted by performing a NotI Pst I double digestion, or from the pPT vector by performing a MunI HindIII double digestion, to produce the appropriate fragment for each clone, provided the cDNA sequence does not contain any of the corresponding restriction sites within its sequence.

Alternatively, other restriction enzymes of the multicloning site of the vector may be used to

5 recover the desired insert as indicated by the manufacturer.

Pools of cells containing certain cDNAs of the invention, from which the cells containing a particular polynucleotide is obtainable, have also been deposited with the American Tissue Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, United States. These cDNA clones have been transfected into separate bacterial cells (E-coli) for these composite

10 deposits.

Bacterial cells containing a particular clone can be obtained from the composite deposit as follows:

An oligonucleotide probe or probes should be designed to the sequence that is known for that particular clone. This sequence can be derived from the sequences provided herein, or from a
15 combination of those sequences. The design of the oligonucleotide probe should preferably follow these parameters:

(a) It should be designed to an area of the sequence which has the fewest ambiguous bases ("N's"), if any;

(b) Preferably, the probe is designed to have a T_m of approximately 80 degrees Celsius
20 (assuming 2 degrees for each A or T and 4 degrees for each G or C). However, probes having melting temperatures between 40 degrees Celsius and 80 degrees Celsius may also be used provided that specificity is not lost.

The oligonucleotide should preferably be labeled with gamma[32 P]ATP (specific activity 6000 Ci/mmol) and T4 polynucleotide kinase using commonly employed techniques for labeling
25 oligonucleotides. Other labeling techniques can also be used. Unincorporated label should preferably be removed by gel filtration chromatography or other established methods. The amount of radioactivity incorporated into the probe should be quantified by measurement in a scintillation counter. Preferably, specific activity of the resulting probe should be approximately 4×10^6 dpm/pmol.

30 The bacterial culture containing the pool of full-length clones should preferably be thawed and 100 μ l of the stock used to inoculate a sterile culture flask containing 25 ml of sterile L-broth containing ampicillin at 100 μ g/ml. The culture should preferably be grown to saturation at 37 degrees Celsius, and the saturated culture should preferably be diluted in fresh L-broth. Aliquots of these dilutions should preferably be plated to determine the dilution and volume which will yield
35 approximately 5000 distinct and well-separated colonies on solid bacteriological media containing L-broth containing ampicillin at 100 μ g/ml and agar at 1.5% in a 150 mm petri dish when grown overnight at 37 degrees Celsius. Other known methods of obtaining distinct, well-separated colonies can also be employed.

Standard colony hybridization procedures should then be used to transfer the colonies to nitrocellulose filters and lyse, denature and bake them.

The filter is then preferably incubated at 65 degrees Celsius for 1 hour with gentle agitation in 6X SSC (20X stock is 175.3 g NaCl/liter, 88.2 g Na citrate/liter, adjusted to pH 7.0 with NaOH) containing 0.5% SDS, 100 pg/ml of yeast RNA, and 10 mM EDTA (approximately 10 ml per 150 mm filter). Preferably, the probe is then added to the hybridization mix at a concentration greater than or equal to 1×10^6 dpm/ml. The filter is then preferably incubated at 65 degrees Celsius with gentle agitation overnight. The filter is then preferably washed in 500 ml of 2X SSC/0.1% SDS at room temperature with gentle shaking for 15 minutes. A third wash with 0.1X SSC/0.5% SDS at 65 degrees Celsius for 30 minutes to 1 hour is optional. The filter is then preferably dried and subjected to autoradiography for sufficient time to visualize the positives on the X-ray film. Other known hybridization methods can also be employed.

The positive colonies are picked, grown in culture, and plasmid DNA isolated using standard procedures. The clones can then be verified by restriction analysis, hybridization analysis, or DNA sequencing. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art.

Alternatively, to recover cDNA inserts from the pool of bacteria, a PCR can be performed on plasmid DNA isolated using standard procedures and primers designed at both ends of the cDNA insertion, including primers designed in the multicloning site of the vector. If a specific cDNA of interest is to be recovered, primers may be designed in order to be specific for the 5' end and the 3' end of this cDNA using sequence information available from the appended sequence listing. The PCR product which corresponds to the cDNA of interest can then be manipulated using standard cloning techniques familiar to those skilled in the art.

Therefore, an object of the invention is an isolated, purified, or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of cDNA inserts of the deposited clone pool. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant GENSET cDNAs consisting of, consisting essentially of, or comprising a nucleotide sequence selected from the group consisting of cDNA inserts of the deposited clone pool.

30 cDNA sequences of the invention

Another object of the invention is a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784, complementary sequences thereto, and fragments thereof. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant GENSET cDNAs consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID NOs:1-169, 339-455, 561-784.

Accordingly, the coding sequence (CDS) or open reading frame (ORF) of each cDNA of the invention refers to the nucleotide sequence beginning with the first nucleotide of the start codon

and ending with the last nucleotide of the stop codon. Similarly, the 5' untranslated region (or 5'UTR) of each cDNA of the invention refers to the nucleotide sequence starting at nucleotide 1 and ending at the nucleotide immediately 5' to the first nucleotide of the start codon. The 3' untranslated region (or 3'UTR) of each cDNA of the invention refers to the nucleotide sequence starting at the nucleotide immediately 3' to the last nucleotide of the stop codon and ending at the last nucleotide of the cDNA.

Untranslated regions

In addition, the invention concerns a purified, isolated, and recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of the 5'UTRs of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool, sequences complementary thereto, and allelic variants thereof. The invention also concerns a purified, isolated, and/or recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of the 3'UTRs of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool, sequences complementary thereto, and allelic variants thereof.

These polynucleotides may be used to detect the presence of GENSET mRNA species in a biological sample using either hybridization or RT-PCR techniques well known to those skilled in the art.

In addition, these polynucleotides may be used as regulatory molecules able to affect the processing and maturation of any polynucleotide including them (either a GENSET polynucleotide or an heterologous polynucleotide), preferably the localization, stability and/or translation of said polynucleotide including them (for a review on UTRs see Decker and Parker, 1995, Derrigo *et al.*, 2000). In particular, 3'UTRs may be used in order to control the stability of heterologous mRNAs in recombinant vectors using any methods known to those skilled in the art including Makrides ((1999) Protein Expr Purif 1999 Nov;17(2):183-202), US Patents 5,925,564; 5,807,707 and 5,756,264, which disclosures are hereby incorporated by reference in their entireties.

Coding sequences

Another object of the invention is an isolated, purified or recombinant polynucleotide comprising the coding sequence of a sequence selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784, clone inserts of the deposited clone pool, and variants thereof.

A further object of the invention is an isolated, purified or recombinant polynucleotide encoding a polypeptide comprising a sequence selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918 and allelic variants thereof. Another object of the invention is an isolated, purified or recombinant polynucleotide encoding a polypeptide comprising a sequence selected from the group consisting of polypeptides encoded by cDNA inserts of the deposited clone pool and allelic variants thereof.

It will be appreciated that should the extent of the coding sequence differ from that indicated in the appended sequence listing as a result of a sequencing error, reverse transcription or amplification error, mRNA splicing, post-translational modification of the encoded protein, enzymatic cleavage of the encoded protein, or other biological factors, one skilled in the art would be readily able to identify the extent of the coding sequences in the sequences of SEQ ID NOs:1-169, 339-455, 561-784. Accordingly, the scope of any claims herein relating to nucleic acids containing the coding sequence of one of SEQ ID NOs:1-169, 339-455, 561-784 is not to be construed as excluding any readily identifiable variations from or equivalents to the coding sequences described in the appended sequence listing. Equivalents includes any alterations in a nucleotide coding sequence that does not result in an amino acid change, or that results in a conservative amino acid substitution, as defined below, in the polypeptide encoded by the nucleotide sequence. Similarly, should the extent of the polypeptides differ from those indicated in the appended sequence listing as a result of any of the preceding factors, the scope of claims relating to polypeptides comprising the amino acid sequence of the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918 is not to be construed as excluding any readily identifiable variations from or equivalents to the sequences described in the appended sequence listing.

The above-disclosed polynucleotides that contain the coding sequence of the GENSET genes may be expressed in a desired host cell or a desired host organism, when this polynucleotide is placed under the control of suitable expression signals. The expression signals may be either the expression signals contained in the regulatory regions in the GENSET genes of the invention or, in contrast, the signals may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression and/or amplification.

Further included in the present invention are polynucleotides encoding the polypeptides of the present invention that are fused in frame to the coding sequences for additional heterologous amino acid sequences. Also included in the present invention are nucleic acids encoding polypeptides of the present invention together with additional, non-coding sequences, including, but not limited to, non-coding 5' and 3' sequences, vector sequence, sequences used for purification, probing, or priming. For example, heterologous sequences include transcribed, untranslated sequences that may play a role in transcription and mRNA processing, such as ribosome binding and stability of mRNA. The heterologous sequences may alternatively comprise additional coding sequences that provide additional functionalities. Thus, a nucleotide sequence encoding a polypeptide may be fused to a tag sequence, such as a sequence encoding a peptide that facilitates purification or detection of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the tag amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN), or in any of a number of additional, commercially available vectors. For instance, hexa-histidine provides for the convenient purification of the fusion protein (see, Gentz *et al.*, 1989, Proc Natl Acad Sci U S A Feb;86(3):821-4, the disclosure of which is incorporated by reference in its entirety). The "HA" tag is another peptide useful for purification which corresponds

to an epitope derived from the influenza hemagglutinin protein (see, Wilson *et al.*, 1984, Cell Jul;37(3):767-78, the disclosure of which is incorporated by reference in its entirety). As discussed below, other such fusion proteins include a GENSET polypeptide fused to Fc at the N- or C-terminus.

- 5 Suitable recombinant vectors that contain a polynucleotide such as described herein are disclosed elsewhere in the specification. Expression vectors encoding GENSET polypeptides or fragments thereof are described in the section entitled "Preparation of the polypeptides".

Regulatory sequences of the invention

- As mentioned, the genomic sequence of GENSET genes contain regulatory sequences in
10 the non-coding 5'-flanking region and possibly in the non-coding 3'-flanking region that border the GENSET polypeptide coding regions containing the exons of these genes.

Polynucleotides derived from GENSET polynucleotide 5' and 3' regulatory regions are useful in order to detect the presence of at least a copy of a genomic nucleotide sequence of the GENSET gene or a fragment thereof in a test sample.

15 Preferred regulatory sequences

Polynucleotides carrying the regulatory elements located at the 5' end and at the 3' end of GENSET polypeptide coding regions may be advantageously used to control, e.g., the transcriptional and translational activity of a heterologous polynucleotide of interest.

- Thus, the present invention also concerns a purified or isolated nucleic acid comprising a
20 polynucleotide which is selected from the group consisting of the 5' and 3' GENSET polynucleotide regulatory regions, sequences complementary thereto, regulatory active fragments and variants thereof. The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of GENSET polynucleotide 5' and 3' regulatory regions, advantageously 99 %
25 nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group consisting of GENSET polynucleotide 5' and 3' regulatory regions, sequences complementary thereto, variants and regulatory active fragments thereof.

- Another object of the invention consists of purified, isolated or recombinant nucleic acids
30 comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide selected from the group consisting of the nucleotide sequences of GENSET polynucleotide 5' and 3' regulatory regions, sequences complementary thereto, variants and regulatory active fragments thereof.

- Preferred fragments of 5' regulatory regions have a length of about 1500 or 1000
35 nucleotides, preferably of about 500 nucleotides, more preferably about 400 nucleotides, even more preferably 300 nucleotides and most preferably about 200 nucleotides.

Preferred fragments of 3' regulatory regions are at least 20, 50, 100, 150, 200, 300 or 400 bases in length.

“Regulatory active” polynucleotide derivatives of the 5' or 3' regulatory region are polynucleotides comprising or alternatively consisting of a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host. It could act either as an enhancer or as a repressor. For the purpose of the invention, a nucleic acid or polynucleotide is “functional” as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are “operably linked” to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide.

The regulatory polynucleotides of the invention may be prepared from the nucleotide sequence of GENSET genomic or cDNA sequence, for example, by cleavage using suitable restriction enzymes, or by PCR. The regulatory polynucleotides may also be prepared by digestion of a GENSET gene-containing genomic clone by an exonuclease enzyme, such as Bal31 (Wabiko *et al.*, DNA 5(4):305-14 (1986), the disclosure of which is incorporated by reference in its entirety). These regulatory polynucleotides can also be prepared by nucleic acid chemical synthesis, as described elsewhere in the specification.

The regulatory polynucleotides according to the invention may be part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism. The recombinant expression vectors according to the invention are described elsewhere in the specification.

Preferred 5'-regulatory polynucleotides of the invention include 5'-UTRs of GENSET cDNAs, or regulatory active fragments or variants thereof. More preferred 5'-regulatory polynucleotides of the invention include sequences selected from the group consisting of 5'-UTRs of sequences of SEQ ID NOs:1-169, 339-455, 561-784, 5'-UTRs of clone inserts of the deposited clone pool, regulatory active fragments and variants thereof.

Preferred 3'-regulatory polynucleotide of the invention include 3'-UTRs of GENSET cDNAs, or regulatory active fragments or variants thereof. More preferred 3'-regulatory polynucleotides of the invention include sequences selected from the group consisting of 3'-UTRs of sequences of SEQ ID NOs:1-169, 339-455, 561-784, 3'-UTRs of clone inserts of the deposited clone pool, regulatory active fragments and variants thereof.

A further object of the invention consists of a purified or isolated nucleic acid comprising:
a) a polynucleotide comprising a 5' regulatory nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence comprising a polynucleotide of a GENSET polynucleotide 5' regulatory region or a complementary sequence thereto;

(ii) a nucleotide sequence comprising a polynucleotide having at least 95% of nucleotide identity with the nucleotide sequence of a GENSET polynucleotide 5' regulatory region or a complementary sequence thereto;

(iii) a nucleotide sequence comprising a polynucleotide that hybridizes under stringent hybridization conditions with the nucleotide sequence of a GENSET polynucleotide 5' regulatory region or a complementary sequence thereto; and

(iv) a regulatory active fragment or variant of the polynucleotides in (i), (ii) and (iii);

b) a nucleic acid molecule encoding a desired polypeptide or a nucleic acid molecule of interest, wherein said nucleic acid molecule is operably linked to the polynucleotide defined in (a);
10 and

c) optionally, a polynucleotide comprising a 3'- regulatory polynucleotide, preferably a 3'- regulatory polynucleotide of a GENSET gene.

In a specific embodiment, the nucleic acid defined above includes the 5'-UTR of a GENSET cDNA, or a regulatory active fragment or variant thereof.

15 In a second specific embodiment, the nucleic acid defined above includes the 3'-UTR of a GENSET cDNA, or a regulatory active fragment or variant thereof.

The regulatory polynucleotide of the 5' regulatory region, or its regulatory active fragments or variants, is operably linked at the 5'-end of the nucleic acid molecule encoding the desired polypeptide or nucleic acid molecule of interest.

20 The regulatory polynucleotide of the 3' regulatory region, or its regulatory active fragments or variants, is advantageously operably linked at the 3'-end of the nucleic acid molecule encoding the desired polypeptide or nucleic acid molecule of interest.

The desired polypeptide encoded by the above-described nucleic acid may be of various nature or origin, encompassing proteins of prokaryotic viral or eukaryotic origin. Among the polypeptides expressed under the control of a GENSET polynucleotide regulatory region include
25 bacterial, fungal or viral antigens. Also encompassed are eukaryotic proteins such as intracellular proteins, such as "house-keeping" proteins, membrane-bound proteins, such as mitochondrial membrane-bound proteins and cell surface receptors, and secreted proteins such as endogenous mediators such as cytokines. The desired polypeptide may be an heterologous polypeptide or a
30 GENSET polypeptide, especially a protein with an amino acid sequence selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918, fragments and variants thereof.

The desired nucleic acids encoded by the above-described polynucleotides, usually an RNA molecule, may be complementary to a desired coding polynucleotide, for example to a GENSET
35 coding sequence, and thus useful as an antisense polynucleotide. Such a polynucleotide may be included in a recombinant expression vector in order to express the desired polypeptide or the desired nucleic acid in host cell or in a host organism. Suitable recombinant vectors that contain a polynucleotide such as described herein are disclosed elsewhere in the specification.

Polynucleotide variants

The invention also relates to variants of the polynucleotides described herein and fragments thereof. “Variants” of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical. The present invention encompasses both allelic variants and degenerate variants.

Examples of variant sequences of polynucleotides of the invention are given in the appended sequence listing. Specifically, Table I includes sequences for which a plurality of closely related sequences, e.g. variants, are provided.

10 Allelic variants

A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. By an “allelic variant” is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism (see Lewin, 1990), the disclosure of which is incorporated by reference in its entirety. Diploid organisms may be homozygous or heterozygous for an allelic form. Non-naturally occurring variants of the polynucleotide may be made by art-known mutagenesis techniques, including those applied to polynucleotides, cells or organisms. See, for example, Table I, which includes sequences for which a plurality of closely related sequences, e.g. allelic variants of a single gene, are provided.

20 Degenerate variant

In addition to the isolated polynucleotides of the present invention, and fragments thereof, the invention further includes polynucleotides which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode a GENSET polypeptide of the present invention. These polynucleotide variants are referred to as “degenerate variants” throughout the instant application. That is, all possible polynucleotide sequences that encode the GENSET polypeptides of the present invention are contemplated. This includes the genetic code and species-specific codon preferences known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above, for instance, to optimize codon expression for a particular host (e.g., change codons in the human mRNA to those preferred by other mammalian or bacterial host cells).

Nucleotide changes present in a variant polynucleotide may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. In the context of the present invention, preferred embodiments

are those in which the polynucleotide variants encode polypeptides which retain substantially the same biological properties or activities as the GENSET protein. More preferred polynucleotide variants are those containing conservative substitutions.

Similar polynucleotides

5 Other embodiments of the present invention provide a purified, isolated or recombinant polynucleotide which is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a polynucleotide selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and the clone inserts of the deposited clone pool. The above polynucleotides are included regardless of whether they encode a polypeptide having a GENSET biological activity. This is
10 because even where a particular nucleic acid molecule does not encode a polypeptide having activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having GENSET activity include, inter alia, isolating a GENSET gene or allelic variants thereof from a DNA library, and detecting GENSET mRNA expression in
15 biological samples suspected of containing GENSET mRNA or DNA, e.g., by Northern Blot or PCR analysis.

The present invention is further directed to polynucleotides having sequences at least 50%. 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% identity to a polynucleotide selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and clone inserts of the
20 deposited clone pool, where said polynucleotide do, in fact, encode a polypeptide having a GENSET biological activity. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides at least 50%. 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% identical to a polynucleotide selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and clone inserts of the
25 deposited clone pool will encode a polypeptide having biological activity. In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having biological activity. This is because the
30 skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly affect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below. By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence
35 except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the GENSET polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted, inserted,

or substituted with another nucleotide. The query sequence may be an entire sequence selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool, or the ORF (open reading frame) of a polynucleotide sequence selected from said group, or any fragment specified as described herein.

5 Hybridizing Polynucleotides

In another aspect, the invention provides an isolated or purified nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to any polynucleotide of the present invention using any methods known to those skilled in the art including those disclosed herein and in particular in the "To find similar sequences" section. Also
10 contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions, preferably at moderate or low stringency conditions as defined herein. Such hybridizing polynucleotides may be of at least 15, 18, 20, 23, 25, 28, 30, 35, 40, 50, 75, 100, 200, 300, 500 or 1000 nucleotides in length.

Of particular interest are polynucleotides hybridizing to any polynucleotide of the invention
15 and encoding GENSET polypeptides, particularly GENSET polypeptides exhibiting a GENSET biological activity.

Of course, a polynucleotide which hybridizes only to polyA⁺ sequences (such as any 3' terminal polyA⁺ tract of a cDNA shown in the sequence listing), or to a 5' complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a
20 polynucleotide would hybridize to any nucleic acid molecule containing a poly(A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone generated using oligo dT as a primer).

Complementary polynucleotides

The invention further provides isolated nucleic acid molecules having a nucleotide
25 sequence fully complementary to any polynucleotide of the invention. The present invention encompasses a purified, isolated or recombinant polynucleotide having a nucleotide sequence complementary to a sequence selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784, sequences of clone inserts of the deposited clone pool and fragments thereof. Such isolated molecules, particularly DNA molecules, are useful as probes for gene
30 mapping and for identifying GENSET mRNA in a biological sample, for instance, by PCR or Northern blot analysis.

Polynucleotide fragments

The present invention is further directed to polynucleotides encoding portions or fragments of the nucleotide sequences described herein. Uses for the polynucleotide fragments of the present
35 invention include probes, primers, molecular weight markers and for expressing the polypeptide fragments of the present invention. Fragments include portions of polynucleotides selected from

the group consisting of a) the sequences of SEQ ID NOs:1-169, 339-455, 561-784, b) genomic GENSET sequences, c) the polynucleotides encoding a polypeptide selected from the group consisting of the sequences of SEQ ID NOs:170-338, 456-560, 785-918, d) the sequences of clone inserts of the deposited clone pool, and e) the polynucleotides encoding the polypeptides encoded by the clone inserts of the deposited clone pool. Particularly included in the present invention is a purified or isolated polynucleotide comprising at least 8 consecutive bases of a polynucleotide of the present invention. In one aspect of this embodiment, the polynucleotide comprises at least 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 800, 1000, 1500, or 2000 consecutive nucleotides of a polynucleotide of the present invention.

10 In addition to the above preferred polynucleotide sizes, further preferred sub-genuses of polynucleotides comprise at least 8 nucleotides, wherein "at least 8" is defined as any integer between 8 and the integer representing the 3' most nucleotide position as set forth in the sequence listing or elsewhere herein. Further included as preferred polynucleotides of the present invention are polynucleotide fragments at least 8 nucleotides in length, as described above, that are further
15 specified in terms of their 5' and 3' position. The 5' and 3' positions are represented by the position numbers set forth in the appended sequence listing. For allelic, degenerate and other variants, position 1 is defined as the 5' most nucleotide of the ORF, i.e., the nucleotide "A" of the start codon with the remaining nucleotides numbered consecutively. Therefore, every combination of a 5' and 3' nucleotide position that a polynucleotide fragment of the present invention, at least 8 contiguous
20 nucleotides in length, could occupy on a polynucleotide of the invention is included in the invention as an individual species. The polynucleotide fragments specified by 5' and 3' positions can be immediately envisaged and are therefore not individually listed solely for the purpose of not unnecessarily lengthening the specification.

It is noted that the above species of polynucleotide fragments of the present invention may
25 alternatively be described by the formula "a to b"; where "a" equals the 5' most nucleotide position and "b" equals the 3' most nucleotide position of the polynucleotide; and further where "a" equals an integer between 1 and the number of nucleotides of the polynucleotide sequence of the present invention minus 8, and where "b" equals an integer between 9 and the number of nucleotides of the polynucleotide sequence of the present invention; and where "a" is an integer smaller than "b" by at
30 least 8.

Therefore, the present invention encompasses isolated, purified, or recombinant polynucleotides which consist of, consist essentially of, or comprise a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 nucleotides of a sequence selected from the group consisting of the sequences of SEQ ID NOs:1-169, 339-455,
35 561-784 and sequences fully complementary thereto.

Other preferred fragments of the invention are polynucleotides comprising polynucleotides encoding domains of polypeptides. Such fragments may be used to obtain other polynucleotides encoding polypeptides having similar domains using hybridization or RT-PCR techniques. Alternatively, these fragments may be used to express a polypeptide domain which may have a

specific biological property. Thus, another object of the invention is an isolated, purified or recombinant polynucleotide encoding a polypeptide consisting of, consisting essentially of, or comprising a contiguous span of at least 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150 or 200 consecutive amino acids of a sequence selected from the group consisting of the sequences of SEQ ID NOs: 170-338, 456-560, 785-918, to the extent that a contiguous span of these lengths is consistent with the lengths of said selected sequence, where said contiguous span comprises at least 1, 2, 3, 5, or 10 of the amino acid positions of a domain of said selected sequence. The present invention also encompasses isolated, purified or recombinant polynucleotides encoding a polypeptide comprising a contiguous span of at least 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150 or 200 consecutive amino acids of a sequence selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918, to the extent that a contiguous span of these lengths is consistent with the lengths of said selected sequence, where said contiguous span is a domain of said selected sequence. The present invention also encompasses isolated, purified or recombinant polynucleotides encoding a polypeptide comprising a domain of a sequence selected from the group consisting of the sequences of SEQ ID NOs: 170-338, 456-560, 785-918.

The present invention further encompasses any combination of the polynucleotide fragments listed in this section.

Oligonucleotide primers and probes

The present invention also encompasses fragments of GENSET polynucleotides for use as primers and probes. Polynucleotides derived from the GENSET genomic and cDNA sequences are useful in order to detect the presence of at least a copy of a GENSET polynucleotide or fragment, complement, or variant thereof in a test sample.

Structural definition

Any polynucleotide of the invention may be used as a primer or probe. Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of the GENSET genomic sequences, the cDNA sequences and the sequences fully complementary thereto. Another object of the invention is a purified, isolated, or recombinant polynucleotide comprising the nucleotide sequence of a sequence selected from the group consisting of the sequences of SEQ ID NOs: 1-169, 339-455, 561-784, sequences of clone inserts of the deposited clone pool, sequences fully complementary thereto, allelic variants thereof, and fragments thereof. Moreover, preferred probes and primers of the invention include purified, isolated, or recombinant GENSET cDNAs consisting of, consisting essentially of, or comprising the sequences of SEQ ID NOs: 1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool.

Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70,

80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of the sequences of SEQ ID NOs:1-169, 339-455, 561-784 and the sequences fully complementary thereto.

Design of primers and probes

- 5 A probe or a primer according to the invention has between 8 and 1000 nucleotides in length, or is specified to be at least 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 nucleotides in length. More particularly, the length of these probes and primers can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 nucleotides. Shorter probes and primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes and primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art. The formation of stable hybrids depends on the melting temperature (T_m) of the DNA. The T_m depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the probes of the invention usually ranges between 10 and 75%, preferably between 35 and 60%, and more preferably between 40 and 55%.
- 20 For amplification purposes, pairs of primers with approximately the same T_m are preferable. Primers may be designed using the OSP software (Hillier and Green, 1991), the disclosure of which is incorporated by reference in its entirety, based on GC content and melting temperatures of oligonucleotides, or using PC-Rare (<http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html>) based on the octamer frequency disparity method (Griffais *et al.*, 1991), the disclosure of which is incorporated by reference in its entirety. DNA amplification techniques are well known to those skilled in the art. Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli *et al.* (1990) and in Compton (1991), Q-beta amplification as described in European Patent Application No 4544610, strand displacement amplification as described in Walker *et al.* (1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461, the disclosures of which are incorporated by reference in their entireties.
- 35 LCR and Gap LCR are exponential amplification techniques, both depending on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a

first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes, which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227), the disclosure of which is incorporated by reference in its entirety. Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall *et al.*(1994), the disclosures of which are incorporated by reference in its entirety. AGLCR is a modification of GLCR that allows the amplification of RNA.

PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see White (1997), Erlich (1992) and the publication entitled "PCR Methods and Applications" ((1991) Cold Spring Harbor Laboratory Press), the disclosures of which are incorporated by reference in their entirety. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, Tth polymerase or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patent Nos. 4,683,195; 4,683,202; and 4,965,188, the disclosures of which are incorporated herein by reference in their entirety.

Preparation of primers and probes

Primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang *et al.*(1979), the phosphodiester method of Brown *et*

al.(1979), the diethylphosphoramidite method of Beaucage *et al.*(1981) and the solid support method described in EP 0 707 592, which disclosures are hereby incorporated by reference in their entireties.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patent Nos. 5,185,444; 5,034,506 and 5,142,047, which disclosures are hereby incorporated by reference in their entireties. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993, which disclosure is hereby incorporated by reference in its entirety, describes modifications, which can be used to render a probe non-extendable.

Labeling of probes

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (including, ^{32}P , ^{35}S , ^3H , ^{125}I), fluorescent dyes (including, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in the French patent No. FR-7810975 or by Urdea *et al* (1988) or Sanchez-Pescador *et al* (1988), which disclosures are hereby incorporated by reference in their entireties. In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea *et al.* in 1991 or in the European patent No. EP 0 225 807 (Chiron), which disclosures are hereby incorporated by reference in their entireties.

The detectable probe may be single stranded or double stranded and may be made using techniques known in the art, including *in vitro* transcription, nick translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting, dot blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or *in vitro* transcription vectors to facilitate the characterization and expression of the hybridizing nucleic acids in the sample. For example, such techniques may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the detectable probe as described herein.

Immobilization of probes

10 A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

25 The probes of the present invention are useful for a number of purposes. They can notably be used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the GENSET gene or mRNA using other techniques. They may also be used for in situ hybridization.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic beads, non-magnetic beads (including polystyrene beads), membranes (including nitrocellulose strips), plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is

oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

15 Oligonucleotide array

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in GENSET genes, may be used for detecting mutations in the coding or in the non-coding sequences of GENSET genes, and may also be used to determine GENSET gene expression in different contexts such as in different tissues, at different stages of a process (embryo development, disease treatment), and in patients versus healthy individuals as described elsewhere in the application.

As used herein, the term “array” means a one dimensional, two dimensional, or multidimensional arrangement of nucleic acids of sufficient length to permit specific detection of gene expression. For example, the array may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The array may include a GENSET genomic DNA, a GENSET cDNA, sequences complementary thereto or fragments thereof. Preferably, the fragments are at least 12, 15, 18, 20, 25, 30, 35, 40 or 50 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. Even more preferably, the fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be “addressable” where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each

polynucleotides location makes these “addressable” arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent No. 5,143,854; PCT publications WO 90/15070 and 92/10092, which disclosures are hereby incorporated by reference in their entireties. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor *et al.*, 1991), which disclosure is hereby incorporated by reference in its entirety. The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as “Very Large Scale Immobilized Polymer Synthesis” (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854; and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which disclosures are hereby incorporated by reference in their entireties, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256, the disclosures of which are incorporated herein by reference in their entireties.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide of the invention, particularly a probe or primer as described herein. Preferably, the invention concerns an array of nucleic acids comprising at least two polynucleotides of the invention, particularly probes or primers as described herein. Preferably, the invention concerns an array of nucleic acids comprising at least five polynucleotides of the invention, particularly probes or primers as described herein.

A preferred embodiment of the present invention is an array of polynucleotides of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 100 or 500 nucleotides in length which includes at least 1, 2, 5, 10, 15, 20, 35, 50 or 100 sequences selected from the group consisting of the sequences of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool, sequences fully complementary thereto, and fragments thereof.

Methods of making the polynucleotides of the invention

The present invention also comprises methods of making the polynucleotides of the invention, including the polynucleotides of SEQ ID NOs:1-169, 339-455, 561-784, genomic DNA obtainable therefrom, or fragments thereof. These methods comprise sequentially linking together nucleotides to produce the nucleic acids having the preceding sequences. Polynucleotides of the

invention may be synthesized either enzymatically using techniques well known to those skilled in the art including amplification or hybridization-based methods as described herein, or chemically.

A variety of chemical methods of synthesizing nucleic acids are known to those skilled in the art. In many of these methods, synthesis is conducted on a solid support. These included the 3' 5 phosphoramidite methods in which the 3' terminal base of the desired oligonucleotide is immobilized on an insoluble carrier. The nucleotide base to be added is blocked at the 5' hydroxyl and activated at the 3' hydroxyl so as to cause coupling with the immobilized nucleotide base. Deblocking of the new immobilized nucleotide compound and repetition of the cycle will produce the desired polynucleotide. Alternatively, polynucleotides may be prepared as described in U.S. 10 Patent No. 5,049,656, which disclosure is hereby incorporated by reference in its entirety. In some embodiments, several polynucleotides prepared as described above are ligated together to generate longer polynucleotides having a desired sequence.

POLYPEPTIDES OF THE INVENTION

The term " GENSET polypeptides" is used herein to embrace all of the proteins and 15 polypeptides of the present invention. The present invention encompasses GENSET polypeptides, including recombinant, isolated or purified GENSET polypeptides consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID NOs:170-338, 456-560, 785-918 and the polypeptides encoded by human cDNAs contained in the deposited clones. Other objects of the invention are polypeptides encoded by the polynucleotides of the 20 invention as well as fusion polypeptides comprising such polypeptides.

Polypeptide variants

The present invention further provides for GENSET polypeptides encoded by allelic and splice variants, orthologs, and/or species homologues. Procedures known in the art can be used to obtain, allelic variants, splice variants, orthologs, and/or species homologues of polynucleotides 25 encoding by polypeptides of the group consisting of SEQ ID NOs:170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool, using information from the sequences disclosed herein or the clones deposited with the ATCC.

The polypeptides of the present invention also include polypeptides having an amino acid sequence at least 50% identical, more preferably at least 60% identical, and still more preferably 30 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% identical to a polypeptide selected from the group consisting of the sequences of SEQ ID NOs:170-338, 456-560, 785-918 and those encoded by the clone inserts of the deposited clone pool. By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except 35 that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% (5 of 100) of

the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid.

Further polypeptides of the present invention include polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above. By a polypeptide having an amino acid sequence at least, for example, 95% "similar" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is similar (i.e. contains identical or equivalent amino acid residues) to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% similar to a query amino acid sequence, up to 5% (5 of 100) of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another non-equivalent amino acid.

These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence. The query sequence may be an entire amino acid sequence selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918 and those encoded by the clone inserts of the deposited clone pool or any fragment specified as described herein.

The variant polypeptides described herein are included in the present invention regardless of whether they have their normal biological activity. This is because even where a particular polypeptide molecule does not have biological activity, one of skill in the art would still know how to use the polypeptide, for instance, as a vaccine or to generate antibodies. Other uses of the polypeptides of the present invention that do not have GENSET biological activity include, *inter alia*, as epitope tags, in epitope mapping, and as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods known to those of skill in the art. As described below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting GENSET protein expression or as agonists and antagonists capable of enhancing or inhibiting GENSET protein function. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" GENSET protein binding proteins, which are also candidate agonists and antagonists according to the present invention (*see, e.g.,* Fields *et al.* 1989, which disclosure is hereby incorporated by reference in its entirety).

Preparation of the polypeptides of the invention

The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of

these methods. The polypeptides of the present invention are preferably provided in an isolated form, and may be partially or preferably substantially purified.

Consequently, the present invention also comprises methods of making the polypeptides of the invention, particularly polypeptides encoded by the cDNAs of SEQ ID NOs:1-169, 339-455, 5 561-784 or by the clone inserts of the deposited clone pool, genomic DNA obtainable therefrom, or fragments thereof and methods of making the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918 or fragments thereof. The methods comprise sequentially linking together amino acids to produce the nucleic polypeptides having the preceding sequences. In some embodiments, the polypeptides made by these methods are 150 amino acids or less in length. In other embodiments, 10 the polypeptides made by these methods are 120 amino acids or less in length.

Isolation

From natural sources

The GENSET proteins of the invention may be isolated from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured cells, of humans or non-human 15 animals. Methods for extracting and purifying natural proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis. See, for example, "Methods in Enzymology, Academic Press, 1993" for a variety of methods for purifying proteins, which 20 disclosure is hereby incorporated by reference in its entirety. Polypeptides of the invention also can be purified from natural sources using antibodies directed against the polypeptides of the invention, such as those described herein, in methods which are well known in the art of protein purification.

From recombinant sources

Preferably, the GENSET polypeptides of the invention are recombinantly produced using 25 routine expression methods known in the art. The polynucleotide encoding the desired polypeptide is operably linked to a promoter into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems are used in forming recombinant polypeptides. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use.

30 Any GENSET polynucleotide, including those described in SEQ ID NOs:1-169, 339-455, 561-784, those of clone inserts of the deposited clone pool, and allelic variants thereof may be used to express GENSET polypeptides. The nucleic acid encoding the GENSET polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The GENSET insert in the expression vector may comprise the full coding sequence 35 for the GENSET protein or a portion thereof. For example, the GENSET derived insert may encode a polypeptide comprising at least 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150 or 200

consecutive amino acids of a GENSET protein selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool.

Consequently, a further embodiment of the present invention is a method of making a polypeptide comprising a protein selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool, said method comprising the steps of:

a) obtaining a cDNA comprising a sequence selected from the group consisting of i) the sequences SEQ ID NOs:1-169, 339-455, 561-784, ii) the sequences of clone inserts of the deposited clone pool one, iii) sequences encoding one of the polypeptide of SEQ ID NOs:170-338, 456-560, 785-918, and iv) sequences of polynucleotides encoding a polypeptide which is encoded by one of the clone insert of the deposited clone pool;

b) inserting said cDNA in an expression vector such that the cDNA is operably linked to a promoter; and

c) introducing said expression vector into a host cell whereby said host cell produces said polypeptide.

In one aspect of this embodiment, the method further comprises the step of isolating the polypeptide. Another embodiment of the present invention is a polypeptide obtainable by the method described in the preceding paragraph.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained in U.S. Patent No. 5,082,767, which disclosure is hereby incorporated by reference in its entirety.

In one embodiment, the entire coding sequence of a GENSET cDNA and the 3'UTR through the poly A signal of the cDNA is operably linked to a promoter in the expression vector.

Alternatively, if the nucleic acid encoding a portion of the GENSET protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the GENSET cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the GENSET protein or a portion thereof is obtained by PCR from a vector containing a GENSET cDNA selected

from the group consisting of the sequences of SEQ ID NOs:1-169, 339-455, 561-784 and the clone inserts of the deposited clone pool using oligonucleotide primers complementary to the GENSET cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure
5 that the sequence encoding the GENSET protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

In another embodiment, it is often advantageous to add to the recombinant polynucleotide
10 additional nucleotide sequence which codes for secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

As a control, the expression vector lacking a cDNA insert is introduced into host cells or organisms.

15 Transfection of a GENSET expression vector into mouse NTH 3T3 cells is but one embodiment of introducing polynucleotides into host cells. Introduction of a polynucleotide encoding a polypeptide into a host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory
20 manuals, such as Davis *et al.* (1986), which disclosure is hereby incorporated by reference in its entirety. It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

Recombinant cell extracts, or proteins from the culture medium if the expressed polypeptide is secreted, are then prepared and proteins separated by gel electrophoresis. If desired, the proteins
25 may be ammonium sulfate precipitated or separated based on size or charge prior to electrophoresis. The proteins present are detected using techniques such as Coomassie or silver staining or using antibodies against the protein encoded by the GENSET cDNA of interest. Coomassie and silver staining techniques are familiar to those skilled in the art.

Proteins from the host cells or organisms containing an expression vector which contains
30 the GENSET cDNA or a fragment thereof are compared to those from the control cells or organism. The presence of a band from the cells containing the expression vector which is absent in control cells indicates that the GENSET cDNA is expressed. Generally, the band corresponding to the protein encoded by the GENSET cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the cDNA. However, the band may have a
35 mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, the GENSET polypeptide to be expressed may also be a product of transgenic animals, i.e., as a component of the milk of transgenic cows, goats, pigs or sheep which

are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein of interest.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including differential extraction, ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. See, for example, "Methods in Enzymology", *supra* for a variety of methods for purifying proteins. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. A recombinantly produced version of a GENSET polypeptide can be substantially purified using techniques described herein or otherwise known in the art, such as, for example, by the one-step method described in Smith and Johnson (1988), which disclosure is hereby incorporated by reference in its entirety. Polypeptides of the invention also can be purified from recombinant sources using antibodies directed against the polypeptides of the invention, such as those described herein, in methods which are well known in the art of protein purification.

Preferably, the recombinantly expressed GENSET polypeptide is purified using standard immunochromatography techniques such as the one described in the section entitled "Immunoaffinity Chromatography". In such procedures, a solution containing the protein of interest, such as the culture medium or a cell extract, is applied to a column having antibodies against the protein attached to the chromatography matrix. The recombinant protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the GENSET cDNA sequence or fragment thereof may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the coding sequence of the GENSET cDNA or fragment thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera may be beta-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to beta-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the beta-globin gene or the nickel binding polypeptide and the GENSET cDNA or fragment thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

One useful expression vector for generating beta-globin chimerics is pSG5 (Stratagene), which encodes rabbit beta-globin. Intron II of the rabbit beta-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega.

Polypeptide may additionally be produced from the construct using *in vitro* translation systems such as the *In vitro* Express™ Translation Kit (Stratagene).

Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

From chemical synthesis

In addition, polypeptides of the invention, especially short protein fragments, can be chemically synthesized using techniques known in the art (*See, e.g.*, Creighton, 1983; and Hunkapiller *et al.*, 1984), which disclosures are hereby incorporated by reference in their entireties. For example, a polypeptide corresponding to a fragment of a polypeptide sequence of the invention can be synthesized by use of a peptide synthesizer. A variety of methods of making polypeptides are known to those skilled in the art, including methods in which the carboxyl terminal amino acid is bound to polyvinyl benzene or another suitable resin. The amino acid to be added possesses blocking groups on its amino moiety and any side chain reactive groups so that only its carboxyl moiety can react. The carboxyl group is activated with carbodiimide or another activating agent and allowed to couple to the immobilized amino acid. After removal of the blocking group, the cycle is repeated to generate a polypeptide having the desired sequence. Alternatively, the methods described in U.S. Patent No. 5,049,656, which disclosure is hereby incorporated by reference in its entirety, may be used.

Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the polypeptide sequence. Non-classical amino acids include, but are not limited to, to the D-isomers of the common amino acids, 2,4-diaminobutyric acid, α-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, γ-Abu, ε-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, fluoroamino acids, designer amino acids such as β-methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

35 Modifications

The invention encompasses polypeptides which are differentially modified during or after translation, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known

protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including but not limited, to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH₄; acetylation, formylation, oxidation, reduction; 5 metabolic synthesis in the presence of tunicamycin; etc.

Additional post-translational modifications encompassed by the invention include, for example, e.g., N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends), attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine 10 residue as a result of prokaryotic host cell expression. The polypeptides may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the protein.

Also provided by the invention are chemically modified derivatives of the polypeptides of the invention which may provide additional advantages such as increased solubility, stability and 15 circulating time of the polypeptide, or decreased immunogenicity. See U.S. Patent No: 4,179,337. The chemical moieties for derivatization may be selected See U.S. Patent No: 4,179,337, which disclosure is hereby incorporated by reference in its entirety. The chemical moieties for derivatization may be selected from water soluble polymers such as polyethylene glycol, ethylene glycol/propylene glycol copolymers, carboxymethylcellulose, dextran, polyvinyl alcohol and the 20 like. The polypeptides may be modified at random positions within the molecule, or at predetermined positions within the molecule and may include one, two, three or more attached chemical moieties.

The polymer may be of any molecular weight, and may be branched or unbranched. For polyethylene glycol, the preferred molecular weight is between about 1 kDa and about 100 kDa (the 25 term "about" indicating that in preparations of polyethylene glycol, some molecules will weigh more, some less, than the stated molecular weight) for ease in handling and manufacturing. Other sizes may be used, depending on the desired therapeutic profile (e.g., the duration of sustained release desired, the effects, if any on biological activity, the ease in handling, the degree or lack of antigenicity and other known effects of the polyethylene glycol to a therapeutic protein or analog).

30 The polyethylene glycol molecules (or other chemical moieties) should be attached to the protein with consideration of effects on functional or antigenic domains of the protein. There are a number of attachment methods available to those skilled in the art, e.g., EP 0 401 384, (coupling PEG to G-CSF), and Malik *et al.* (1992) (reporting pegylation of GM-CSF using tresyl chloride), which disclosures are hereby incorporated by reference in their entireties. For example, 35 polyethylene glycol may be covalently bound through amino acid residues via a reactive group, such as, a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group may include lysine residues and the N-terminal amino acid residues; those having a free carboxyl group may include aspartic acid residues glutamic acid residues and the C-terminal amino acid

residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecules. Preferred for therapeutic purposes is attachment at an amino group, such as attachment at the N-terminus or lysine group.

One may specifically desire proteins chemically modified at the N-terminus. Using polyethylene glycol as an illustration of the present composition, one may select from a variety of polyethylene glycol molecules (by molecular weight, branching, etc.), the proportion of polyethylene glycol molecules to protein (polypeptide) molecules in the reaction mix, the type of pegylation reaction to be performed, and the method of obtaining the selected N-terminally pegylated protein. The method of obtaining the N-terminally pegylated preparation (i.e., separating this moiety from other monopegylated moieties if necessary) may be by purification of the N-terminally pegylated material from a population of pegylated protein molecules. Selective proteins chemically modified at the N-terminus modification may be accomplished by reductive alkylation, which exploits differential reactivity of different types of primary amino groups (lysine versus the N-terminal) available for derivatization in a particular protein. Under the appropriate reaction conditions, substantially selective derivatization of the protein at the N-terminus with a carbonyl group containing polymer is achieved.

Multimerization

The polypeptides of the invention may be in monomers or multimers (i.e., dimers, trimers, tetramers and higher multimers). Accordingly, the present invention relates to monomers and multimers of the polypeptides of the invention, their preparation, and compositions containing them. In specific embodiments, the polypeptides of the invention are monomers, dimers, trimers or tetramers. In additional embodiments, the multimers of the invention are at least dimers, at least trimers, or at least tetramers.

Multimers encompassed by the invention may be homomers or heteromers. As used herein, the term "homomer", refers to a multimer containing only polypeptides corresponding to the amino acid sequences of SEQ ID NOs:170-338, 456-560, 785-918 or encoded by the clone inserts of the deposited clone pool (including fragments, variants, splice variants, and fusion proteins, corresponding to these polypeptides as described herein). These homomers may contain polypeptides having identical or different amino acid sequences. In a specific embodiment, a homomer of the invention is a multimer containing only polypeptides having an identical amino acid sequence. In another specific embodiment, a homomer of the invention is a multimer containing polypeptides having different amino acid sequences. In specific embodiments, the multimer of the invention is a homodimer (e.g., containing polypeptides having identical or different amino acid sequences) or a homotrimer (e.g., containing polypeptides having identical and/or different amino acid sequences). In additional embodiments, the homomeric multimer of the invention is at least a homodimer, at least a homotrimer, or at least a homotetramer.

As used herein, the term "heteromer" refers to a multimer containing one or more heterologous polypeptides (i.e., polypeptides of different proteins) in addition to the polypeptides of

the invention. In a specific embodiment, the multimer of the invention is a heterodimer, a heterotrimer, or a heterotetramer. In additional embodiments, the heteromeric multimer of the invention is at least a heterodimer, at least a heterotrimer, or at least a heterotetramer.

Multimers of the invention may be the result of hydrophobic, hydrophilic, ionic and/or covalent associations and/or may be indirectly linked, by for example, liposome formation. Thus, in one embodiment, multimers of the invention, such as, for example, homodimers or homotrimers, are formed when polypeptides of the invention contact one another in solution. In another embodiment, heteromultimers of the invention, such as, for example, heterotrimers or heterotetramers, are formed when polypeptides of the invention contact antibodies to the polypeptides of the invention (including antibodies to the heterologous polypeptide sequence in a fusion protein of the invention) in solution. In other embodiments, multimers of the invention are formed by covalent associations with and/or between the polypeptides of the invention. Such covalent associations may involve one or more amino acid residues contained in the polypeptide sequence (e.g., that recited in the sequence listing, or contained in the polypeptide encoded by a deposited clone). In one instance, the covalent associations are cross-linking between cysteine residues located within the polypeptide sequences, which interact in the native (i.e., naturally occurring) polypeptide. In another instance, the covalent associations are the consequence of chemical or recombinant manipulation. Alternatively, such covalent associations may involve one or more amino acid residues contained in the heterologous polypeptide sequence in a fusion protein of the invention.

In one example, covalent associations are between the heterologous sequence contained in a fusion protein of the invention (see, e.g., US Patent Number 5,478,925, which disclosure is hereby incorporated by reference in its entirety). In a specific example, the covalent associations are between the heterologous sequence contained in an Fc fusion protein of the invention (as described herein). In another specific example, covalent associations of fusion proteins of the invention are between heterologous polypeptide sequence from another protein that is capable of forming covalently associated multimers, such as for example, osteopontin (see, e.g., International Publication No: WO 98/49305, the contents of which are herein incorporated by reference in its entirety). In another embodiment, two or more polypeptides of the invention are joined through peptide linkers. Examples include those peptide linkers described in U.S. Pat. No. 5,073,627 (hereby incorporated by reference). Proteins comprising multiple polypeptides of the invention separated by peptide linkers may be produced using conventional recombinant DNA technology.

Another method for preparing multimer polypeptides of the invention involves the use of polypeptides of the invention fused to a leucine zipper or isoleucine zipper polypeptide sequence. Leucine zipper and isoleucine zipper domains are polypeptides that promote multimerization of the proteins in which they are found. Leucine zippers were originally identified in several DNA-binding proteins, and have since been found in a variety of different proteins (Landschulz *et al.*, 1988). Among the known leucine zippers are naturally occurring peptides and derivatives thereof that dimerize or trimerize. Examples of leucine zipper domains suitable for producing

soluble multimeric proteins of the invention are those described in PCT application WO 94/10308, hereby incorporated by reference. Recombinant fusion proteins comprising a polypeptide of the invention fused to a polypeptide sequence that dimerizes or trimerizes in solution are expressed in suitable host cells, and the resulting soluble multimeric fusion protein is recovered from the culture supernatant using techniques known in the art.

Trimeric polypeptides of the invention may offer the advantage of enhanced biological activity. Preferred leucine zipper moieties and isoleucine moieties are those that preferentially form trimers. One example is a leucine zipper derived from lung surfactant protein D (SPD), as described in Hoppe *et al.* (1994) and in U.S. patent application Ser. No. 08/446,922, which disclosure is hereby incorporated by reference in its entirety. Other peptides derived from naturally occurring trimeric proteins may be employed in preparing trimeric polypeptides of the invention. In another example, proteins of the invention are associated by interactions between Flag® polypeptide sequence contained in fusion proteins of the invention containing Flag® polypeptide sequence. In a further embodiment, associations proteins of the invention are associated by interactions between heterologous polypeptide sequence contained in Flag® fusion proteins of the invention and anti Flag® antibody.

The multimers of the invention may be generated using chemical techniques known in the art. For example, polypeptides desired to be contained in the multimers of the invention may be chemically cross-linked using linker molecules and linker molecule length optimization techniques known in the art (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Additionally, multimers of the invention may be generated using techniques known in the art to form one or more inter-molecule cross-links between the cysteine residues located within the sequence of the polypeptides desired to be contained in the multimer (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Further, polypeptides of the invention may be routinely modified by the addition of cysteine or biotin to the C terminus or N-terminus of the polypeptide and techniques known in the art may be applied to generate multimers containing one or more of these modified polypeptides (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). Additionally, other techniques known in the art may be applied to generate liposomes containing the polypeptide components desired to be contained in the multimer of the invention (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety).

Alternatively, multimers of the invention may be generated using genetic engineering techniques known in the art. In one embodiment, polypeptides contained in multimers of the invention are produced recombinantly using fusion protein technology described herein or otherwise known in the art (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). In a specific embodiment, polynucleotides coding for a homodimer of the invention are generated by ligating a polynucleotide sequence encoding a polypeptide of the invention to a sequence encoding a linker polypeptide and then further to a synthetic polynucleotide encoding the translated product of the polypeptide in the reverse orientation from the original

C-terminus to the N-terminus (lacking the leader sequence) (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety). In another embodiment, recombinant techniques described herein or otherwise known in the art are applied to generate recombinant polypeptides of the invention which contain a transmembrane domain (or hydrophobic or signal peptide) and which can be incorporated by membrane reconstitution techniques into liposomes (see, e.g., US Patent Number 5,478,925, which is herein incorporated by reference in its entirety).

Mutated polypeptides

To improve or alter the characteristics of GENSET polypeptides of the present invention, protein engineering may be employed. Recombinant DNA technology known to those skilled in the art can be used to create novel mutant proteins or muteins including single or multiple amino acid substitutions, deletions, additions, or fusion proteins. Such modified polypeptides can show, e.g., increased/decreased biological activity or increased/decreased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions. Further, the polypeptides of the present invention may be produced as multimers including dimers, trimers and tetramers. Multimerization may be facilitated by linkers or recombinantly through heterologous polypeptides such as Fc regions.

N- and C-terminal deletions

It is known in the art that one or more amino acids may be deleted from the N-terminus or C-terminus without substantial loss of biological function. For instance, Ron *et al.* (1993), reported modified KGF proteins that had heparin binding activity even if 3, 8, or 27 N-terminal amino acid residues were missing. Accordingly, the present invention provides polypeptides having one or more residues deleted from the amino terminus of the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918 or that encoded by the clone inserts of the deposited clone pool. Similarly, many examples of biologically functional C-terminal deletion mutants are known. For instance, Interferon gamma shows up to ten times higher activities by deleting 810 amino acid residues from the C-terminus of the protein (See, e.g., Dobeli, *et al.* 1988), which disclosure is hereby incorporated by reference in its entirety. Accordingly, the present invention provides polypeptides having one or more residues deleted from the carboxy terminus of the polypeptides shown of SEQ ID NOs:170-338, 456-560, 785-918 or encoded by the clone inserts of the deposited clone pool. The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini as described below.

Other mutations

Other mutants in addition to N- and C-terminal deletion forms of the protein discussed above are included in the present invention. It also will be recognized by one of ordinary skill in the art that some amino acid sequences of the GENSET polypeptides of the present invention can be varied without significant effect of the structure or function of the protein. If such differences in

sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity. Thus, the invention further includes variations of the GENSET polypeptides which show substantial GENSET polypeptide activity. Such mutants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as to have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided.

There are two main approaches for studying the tolerance of an amino acid sequence to change (see, Bowie *et al.* 1994, which disclosure is hereby incorporated by reference in its entirety). The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection.

The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality. These studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The studies indicate which amino acid changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described by Bowie *et al.* (supra) and the references cited therein.

Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Phe; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Thus, the fragment, derivative, analog, or homologue of the polypeptide of the present invention may be, for example: (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code: or (ii) one in which one or more of the amino acid residues includes a substituent group: or (iii) one in which the GENSET polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol): or (iv) one in which the additional amino acids are fused to the above form of the polypeptide, such as an IgG Fc fusion region peptide or leader or secretory sequence or a sequence which is employed for purification of the above form of the polypeptide or a pro-protein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Thus, the GENSET polypeptides of the present invention may include one or more amino acid substitutions, deletions, or additions, either from natural mutations or human manipulation. As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein. The following groups of amino acids generally represent equivalent changes: (1) Ala, Pro, Gly, Glu, Asp, Gln, Asn, Ser, Thr; (2) Cys, Ser, Tyr, Thr; (3) Val, Ile, Leu, Met, Ala, Phe; (4) Lys, Arg, His; (5) Phe, Tyr, Trp, His.

A specific embodiment of a modified GENSET peptide molecule of interest according to the present invention, includes, but is not limited to, a peptide molecule which is resistant to proteolysis, is a peptide in which the -CONH- peptide bond is modified and replaced by a (CH₂NH) reduced bond, a (NHCO) retro inverso bond, a (CH₂-O) methylene-oxy bond, a (CH₂-S) thiomethylene bond, a (CH₂CH₂) carba bond, a (CO-CH₂) cetomethylene bond, a (CHOH-CH₂) hydroxyethylene bond), a (N-N) bound, a E-alcene bond or also a -CH=CH- bond. The invention also encompasses a human GENSET polypeptide or a fragment or a variant thereof in which at least one peptide bond has been modified as described above.

Amino acids in the GENSET proteins of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (*see, e.g.,* Cunningham *et al.* 1989, which disclosure is hereby incorporated by reference in its entirety). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity using assays appropriate for measuring the function of the particular protein. Of special interest are substitutions of charged amino acids with other charged or neutral amino acids which may produce proteins with highly desirable improved characteristics, such as less aggregation. Aggregation may not only reduce activity but also be problematic when preparing pharmaceutical formulations, because aggregates can be immunogenic, (*see, e.g.,* Pinckard *et al.*, 1967; Robbins, *et al.*, 1987; and Cleland, *et al.*, 1993).

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of a GENSET polypeptide having an amino acid sequence which contains at least one conservative amino acid substitution, but not more than 50 conservative amino acid substitutions, not more than 40 conservative amino acid substitutions, not more than 30 conservative amino acid substitutions, and not more than 20 conservative amino acid substitutions. Also provided are polypeptides which comprise the amino acid sequence of a GENSET polypeptide, having at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

Polypeptide fragments

Structural definition

The present invention is further directed to fragments of the amino acid sequences described herein such as the polypeptides of SEQ ID NOs:170-338, 456-560, 785-918 or those encoded by the clone inserts of the deposited clone pool. More specifically, the present invention embodies purified, isolated, and recombinant polypeptides comprising at least 6, preferably at least 8 to 10, more preferably 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 125, 150, 175, 200, 225, 250, 275, or 300 consecutive amino acids of a polypeptide selected from the group consisting of the sequences of SEQ ID NOs:170-338, 456-560, 785-918, the polypeptides encoded by the clone inserts of the deposited clone pool, and other polypeptides of the present invention.

In addition to the above polypeptide fragments, further preferred sub-genuses of polypeptides comprise at least 6 amino acids, wherein "at least 6" is defined as any integer between 6 and the integer representing the C-terminal amino acid of the polypeptide of the present invention including the polypeptide sequences of the sequence listing below. Further included are species of polypeptide fragments at least 6 amino acids in length, as described above, that are further specified in terms of their N-terminal and C-terminal positions. However, included in the present invention as individual species are all polypeptide fragments, at least 6 amino acids in length, as described above, and may be particularly specified by a N-terminal and C-terminal position. That is, every combination of a N-terminal and C-terminal position that a fragment at least 6 contiguous amino acid residues in length could occupy, on any given amino acid sequence of the sequence listing or of the present invention is included in the present invention

The present invention also provides for the exclusion of any fragment species specified by N-terminal and C-terminal positions or of any fragment sub-genus specified by size in amino acid residues as described above. Any number of fragments specified by N-terminal and C-terminal positions or by size in amino acid residues as described above may be excluded as individual species.

The above polypeptide fragments of the present invention can be immediately envisaged using the above description and are therefore not individually listed solely for the purpose of not unnecessarily lengthening the specification. Moreover, the above fragments need not have a GENSET biological activity, although polypeptides having these activities are preferred embodiments of the invention, since they would be useful, for example, in immunoassays, in epitope mapping, epitope tagging, as vaccines, and as molecular weight markers. The above fragments may also be used to generate antibodies to a particular portion of the polypeptide. These antibodies can then be used in immunoassays well known in the art to distinguish between human and non-human cells and tissues or to determine whether cells or tissues in a biological sample are or are not of the same type which express the polypeptides of the present invention.

It is noted that the above species of polypeptide fragments of the present invention may alternatively be described by the formula "a to b"; where "a" equals the N-terminal most amino acid position and "b" equals the C-terminal most amino acid position of the polynucleotide; and further where "a" equals an integer between 1 and the number of amino acids of the polypeptide sequence of the present invention minus 6, and where "b" equals an integer between 7 and the number of amino acids of the polypeptide sequence of the present invention; and where "a" is an integer smaller than "b" by at least 6.

The present invention also provides for the exclusion of any species of polypeptide fragments of the present invention specified by 5' and 3' positions or sub-genuses of polypeptides specified by size in amino acids as described above. Any number of fragments specified by 5' and 3' positions or by size in amino acids, as described above, may be excluded.

Functional definition*Domains*

Preferred polynucleotide fragments of the invention are domains of polypeptides of the invention. Such domains may eventually comprise linear or structural motifs and signatures including, but not limited to, leucine zippers, helix-turn-helix motifs, post-translational modification sites such as glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites. Such domains may present a particular biological activity such as DNA or RNA-binding, secretion of proteins, transcription regulation, enzymatic activity, substrate binding activity, etc.

A domain has a size generally comprised between 3 and 1000 amino acids. In a preferred embodiment, domains comprise a number of amino acids that is any integer between 6 and 200. Domains may be synthesized using any methods known to those skilled in the art, including those disclosed herein, particularly in the section entitled "Preparation of the polypeptides of the invention". Methods for determining the amino acids which make up a domain with a particular biological activity include mutagenesis studies and assays to determine the biological activity to be tested.

Alternatively, the polypeptides of the invention may be scanned for motifs, domains and/or signatures in databases using any computer method known to those skilled in the art. Searchable databases include Prosite (Hofmann *et al.*, 1999; Bucher and Bairoch 1994), Pfam (Sonnhammer *et al.*, 1997; Henikoff *et al.*, 2000; Bateman *et al.*, 2000), Blocks (Henikoff *et al.*, 2000), Print (Attwood *et al.*, 1996), Prodom (Sonnhammer and Kahn, 1994; Corpet *et al.* 2000), Sbase (Pongor *et al.*, 1993; Murvai *et al.*, 2000), Smart (Schultz *et al.*, 1998), Dali/FSSP (Holm and Sander, 1996, 1997 and 1999), HSSP (Sander and Schneider 1991), CATH (Orengo *et al.*, 1997; Pearl *et al.*, 2000), SCOP (Murzin *et al.*, 1995; Lo Conte *et al.*, 2000), COG (Tatusov *et al.*, 1997 and 2000), specific family databases and derivatives thereof (Nevill-Manning *et al.*, 1998; Yona *et al.*, 1999; Attwood *et al.*, 2000), each of which disclosures are hereby incorporated by reference in their entirety. For a review on available databases, see issue 1 of volume 28 of Nucleic Acid Research (2000), which disclosure is hereby incorporated by reference in its entirety.

The domains of the present invention preferably comprises 6 to 200 amino acids (i.e. any integer between 6 and 200, inclusive) of a polypeptide of the present invention. Also, included in the present invention are domain fragments between the integers of 6 and the full length GENSET sequence of the sequence listing. All combinations of sequences between the integers of 6 and the full-length sequence of a GENSET polypeptide are included. The domain fragments may be specified by either the number of contiguous amino acid residues (as a sub-genus) or by specific N-terminal and C-terminal positions (as species) as described above for the polypeptide fragments of

the present invention. Any number of domain fragments of the present invention may also be excluded in the same manner.

Epitopes and Antibody Fusions:

A preferred embodiment of the present invention is directed to epitope-bearing polypeptides and epitope-bearing polypeptide fragments. These epitopes may be "antigenic epitopes" or both an "antigenic epitope" and an "immunogenic epitope". An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response *in vivo* when the polypeptide is the immunogen. On the other hand, a region of polypeptide to which an antibody binds is defined as an "antigenic determinant" or "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes (*See, e.g., Geysen, et al., 1984*), which disclosure is hereby incorporated by reference in its entirety. It is particularly noted that although a particular epitope may not be immunogenic, it is nonetheless useful since antibodies can be made to both immunogenic and antigenic epitopes.

An epitope can comprise as few as 3 amino acids in a spatial conformation, which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more often at least 8-10 such amino acids. In preferred embodiment, antigenic epitopes comprise a number of amino acids that is any integer between 3 and 50. Fragments which function as epitopes may be produced by any conventional means (*See, e.g., Houghten, 1985*), also further described in U.S. Patent No. 4,631,21, which disclosures are hereby incorporated by reference in their entireties. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping, e.g., the Pepscan method described by Geysen *et al.* (1984); PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506, which disclosures are hereby incorporated by reference in their entireties. Another example is the algorithm of Jameson and Wolf, (1988) (said reference incorporated by reference in its entirety). The Jameson-Wolf antigenic analysis, for example, may be performed using the computer program PROTEAN, using default parameters (Version 4.0 Windows, DNASTAR, Inc., 1228 South Park Street Madison, WI).

The epitope-bearing fragments of the present invention preferably comprise 6 to 50 amino acids (i.e. any integer between 6 and 50, inclusive) of a polypeptide of the present invention. Also, included in the present invention are antigenic fragments between the integers of 6 and the full length GENSET sequence of the sequence listing. All combinations of sequences between the integers of 6 and the full-length sequence of a GENSET polypeptide are included. The epitope-bearing fragments may be specified by either the number of contiguous amino acid residues (as a sub-genus) or by specific N-terminal and C-terminal positions (as species) as described above for the polypeptide fragments of the present invention. Any number of epitope-bearing fragments of the present invention may also be excluded in the same manner.

Antigenic epitopes are useful, for example, to raise antibodies, including monoclonal antibodies that specifically bind the epitope (see, Wilson *et al.*, 1984; and Sutcliffe, *et al.*, 1983,

which disclosures are hereby incorporated by reference in their entireties). The antibodies are then used in various techniques such as diagnostic and tissue/cell identification techniques, as described herein, and in purification methods such as immunoaffinity chromatography.

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art (See, Sutcliffe *et al.*, *supra*; Wilson *et al.*, *supra*; Chow *et al.* (1985); and Bittle, *et al.*, (1985), which disclosures are hereby incorporated by reference in their entireties). A preferred immunogenic epitope includes the natural GENSET protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.).

Epitope-bearing polypeptides of the present invention are used to induce antibodies according to methods well known in the art including, but not limited to, *in vivo* immunization, *in vitro* immunization, and phage display methods (See, e.g., Sutcliffe, *et al.*, *supra*; Wilson, *et al.*, *supra*, and Bittle, *et al.*, *supra*). If *in vivo* immunization is used, animals may be immunized with free peptide; however, anti-peptide antibody titer may be boosted by coupling of the peptide to a macromolecular carrier, such as keyhole limpet hemacyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine residues may be coupled to a carrier using a linker such as -maleimidobenzoyl- N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carriers using a more general linking agent such as glutaraldehyde. Animals such as rabbits, rats and mice are immunized with either free or carrier-coupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 µg of peptide or carrier protein and Freund's adjuvant. Several booster injections may be needed, for instance, at intervals of about two weeks, to provide a useful titer of anti-peptide antibody, which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for instance, by adsorption to the peptide on a solid support and elution of the selected antibodies according to methods well known in the art.

As one of skill in the art will appreciate, and discussed above, the polypeptides of the present invention comprising an immunogenic or antigenic epitope can be fused to heterologous polypeptide sequences. For example, the polypeptides of the present invention may be fused with the constant domain of immunoglobulins (IgA, IgE, IgG, IgM), or portions thereof (CH1, CH2, CH3, any combination thereof including both entire domains and portions thereof) resulting in chimeric polypeptides. These fusion proteins facilitate purification, and show an increased half-life *in vivo*. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (See, e.g., EPA 0,394,827; and Traunecker *et al.*, 1988, which disclosures are hereby incorporated by reference in their entireties). Fusion proteins that have a

disulfide-linked dimeric structure due to the IgG portion can also be more efficient in binding and neutralizing other molecules than monomeric polypeptides or fragments thereof alone (*See, e.g.,* Fountoulakis *et al.*, 1995, which disclosure is hereby incorporated by reference in its entirety). Nucleic acids encoding the above epitopes can also be recombined with a gene of interest as an epitope tag to aid in detection and purification of the expressed polypeptide.

Additional fusion proteins of the invention may be generated through the techniques of gene-shuffling, motif-shuffling, exon-shuffling, or codon-shuffling (collectively referred to as "DNA shuffling"). DNA shuffling may be employed to modulate the activities of polypeptides of the present invention thereby effectively generating agonists and antagonists of the polypeptides.

10 See, for example, U.S. Patent Nos.: 5,605,793; 5,811,238; 5,834,252; 5,837,458; and Patten, *et al.*, (1997); Harayama, (1998); Hansson, *et al* (1999); and Lorenzo and Blasco, (1998). (Each of these documents are hereby incorporated by reference). In one embodiment, one or more components, motifs, sections, parts, domains, fragments, etc., of coding polynucleotides of the invention, or the polypeptides encoded thereby may be recombined with one or more components, motifs, sections,

15 parts, domains, fragments, etc. of one or more heterologous molecules.

The present invention further encompasses any combination of the polypeptide fragments listed in this section.

Antibodies

20 Definitions

The present invention further relates to antibodies and T-cell antigen receptors (TCR), which specifically bind the polypeptides, and more specifically, the epitopes of the polypeptides of the present invention. The antibodies of the present invention include IgG (including IgG1, IgG2, IgG3, and IgG4), IgA (including IgA1 and IgA2), IgD, IgE, or IgM, and IgY. The term "antibody"

25 (Ab) refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where a binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. As used herein, the term "antibody" is meant to

30 include whole antibodies, including single-chain whole antibodies, and antigen binding fragments thereof. In a preferred embodiment the antibodies are human antigen binding antibody fragments of the present invention include, but are not limited to, Fab, Fab' F(ab)2 and F(ab')2, Fd, single-chain Fvs (scFv), single-chain antibodies, disulfide-linked Fvs (sdFv) and fragments comprising either a V_L or V_H domain. The antibodies may be from any animal origin including birds and mammals.

35 Preferably, the antibodies are human, murine, rabbit, goat, guinea pig, camel, horse, or chicken.

Antigen-binding antibody fragments, including single-chain antibodies, may comprise the variable region(s) alone or in combination with the entire or partial of the following: hinge region,

CH1, CH2, and CH3 domains. Also included in the invention are any combinations of variable region(s) and hinge region, CH1, CH2, and CH3 domains. The present invention further includes chimeric, humanized, and human monoclonal and polyclonal antibodies, which specifically bind the polypeptides of the present invention. The present invention further includes antibodies that are

5 anti-idiotypic to the antibodies of the present invention.

The antibodies of the present invention may be monospecific, bispecific, and trispecific or have greater multispecificity. Multispecific antibodies may be specific for different epitopes of a polypeptide of the present invention or may be specific for both a polypeptide of the present invention as well as for heterologous compositions, such as a heterologous polypeptide or solid

10 support material. *See, e.g.*, WO 93/17715; WO 92/08802; WO 91/00360; WO 92/05793; Tutt, *et al.* (1991); US Patents 5,573,920, 4,474,893, 5,601,819, 4,714,681, 4,925,648; Kostelny *et al.* (1992), which disclosures are hereby incorporated by reference in their entireties.

Antibodies of the present invention may be described or specified in terms of the epitope(s) or epitope-bearing portion(s) of a polypeptide of the present invention, which are recognized or

15 specifically bound by the antibody. The antibodies may specifically bind a complete protein encoded by a nucleic acid of the present invention, or a fragment thereof. Therefore, the epitope(s) or epitope bearing polypeptide portion(s) may be specified as described herein, *e.g.*, by N-terminal and C-terminal positions, by size in contiguous amino acid residues, or otherwise described herein (including the sequence listing). Antibodies which specifically bind any epitope or polypeptide of

20 the present invention may also be excluded as individual species. Therefore, the present invention includes antibodies that specifically bind specified polypeptides of the present invention, and allows for the exclusion of the same.

Thus, another embodiment of the present invention is a purified or isolated antibody capable of specifically binding to a polypeptide comprising a sequence selected from the group

25 consisting of the sequences of SEQ ID NOs:170-338, 456-560, 785-918 and the sequences of the clone inserts of the deposited clone pool. In one aspect of this embodiment, the antibody is capable of binding to an epitope-containing polypeptide comprising at least 6 consecutive amino acids, preferably at least 8 to 10 consecutive amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 consecutive amino acids of a sequence selected from the group consisting of SEQ ID

30 NOs:170-338, 456-560, 785-918 and sequences of the clone inserts of the deposited clone pool.

Antibodies of the present invention may also be described or specified in terms of their cross-reactivity. Antibodies that do not specifically bind any other analog, ortholog, or homologue of the polypeptides of the present invention are included. Antibodies that do not bind polypeptides with less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less

35 than 65%, less than 60%, less than 55%, and less than 50% identity (as calculated using methods known in the art and described herein, *e.g.*, using FASTDB and the parameters set forth herein) to a polypeptide of the present invention are also included in the present invention. Further included in the present invention are antibodies, which only bind polypeptides encoded by polynucleotides, which hybridize to a polynucleotide of the present invention under stringent hybridization

conditions (as described herein). Antibodies of the present invention may also be described or specified in terms of their binding affinity. Preferred binding affinities include those with a dissociation constant or K_d less than $5 \times 10^{-6}M$, $10^{-6}M$, $5 \times 10^{-7}M$, $10^{-7}M$, $5 \times 10^{-8}M$, $10^{-8}M$, $5 \times 10^{-9}M$, $10^{-9}M$, $5 \times 10^{-10}M$, $10^{-10}M$, $5 \times 10^{-11}M$, $10^{-11}M$, $5 \times 10^{-12}M$, $10^{-12}M$, $5 \times 10^{-13}M$, $10^{-13}M$, $5 \times 10^{-14}M$, $10^{-14}M$, $5 \times 10^{-15}M$, and $10^{-15}M$.

The invention also concerns a purified or isolated antibody capable of specifically binding to a mutated GENSET protein or to a fragment or variant thereof comprising an epitope of the mutated GENSET protein.

10 Preparation of antibodies

The antibodies of the present invention may be prepared by any suitable method known in the art. Some of these methods are described in more detail in the example entitled "Preparation of Antibody Compositions to the GENSET protein". For example, a polypeptide of the present invention or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing "polyclonal antibodies". As used herein, the term "monoclonal antibody" is not limited to antibodies produced through hybridoma technology but it rather refers to an antibody that is derived from a single clone, including eukaryotic, prokaryotic, or phage clone, and not the method by which it is produced. Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display technology.

Hybridoma techniques include those known in the art (*See, e.g., Harlow et al.* 1988; Hammerling, *et al.*, 1981). (Said references incorporated by reference in their entireties). Fab and F(ab')₂ fragments may be produced, for example, from hybridoma-produced antibodies by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments).

Alternatively, antibodies of the present invention can be produced through the application of recombinant DNA technology or through synthetic chemistry using methods known in the art. For example, the antibodies of the present invention can be prepared using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of a phage particle, which carries polynucleotide sequences encoding them. Phage with a desired binding property are selected from a repertoire or combinatorial antibody library (e.g. human or murine) by selecting directly with antigen, typically antigen bound or captured to a solid surface or bead. Phage used in these methods are typically filamentous phage including fd and M13 with Fab, Fv or disulfide stabilized Fv antibody domains recombinantly fused to either the phage gene III or gene VIII protein. Examples of phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman *et al.* (1995); Ames, *et al.* (1995); Kettleborough, *et al.* (1994); Persic, *et al.* (1997); Burton *et al.* (1994); PCT/GB91/01134; WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO

95/20401; and US Patents 5,698,426, 5,223,409, 5,403,484, 5,580,717, 5,427,908, 5,750,753, 5,821,047, 5,571,698, 5,427,908, 5,516,637, 5,780,225, 5,658,727 and 5,733,743 (said references incorporated by reference in their entireties).

As described in the above references, after phage selection, the antibody coding regions
5 from the phage can be isolated and used to generate whole antibodies, including human antibodies, or any other desired antigen binding fragment, and expressed in any desired host including mammalian cells, insect cells, plant cells, yeast, and bacteria. For example, techniques to recombinantly produce Fab, Fab' F(ab)2 and F(ab')2 fragments can also be employed using methods known in the art such as those disclosed in WO 92/22324; Mullinax *et al.* (1992); and
10 Sawai *et al.* (1995); and Better *et al.* (1988) (said references incorporated by reference in their entireties).

Examples of techniques which can be used to produce single-chain Fvs and antibodies include those described in U.S. Patents 4,946,778 and 5,258,498; Huston *et al.* (1991); Shu *et al.* (1993); and Skerra *et al.* (1988), which disclosures are hereby incorporated by reference in their
15 entireties. For some uses, including *in vivo* use of antibodies in humans and *in vitro* detection assays, it may be preferable to use chimeric, humanized, or human antibodies. Methods for producing chimeric antibodies are known in the art. *See e.g.*, Morrison, (1985); Oi *et al.*, (1986); Gillies *et al.* (1989); and US Patent 5,807,715, which disclosures are hereby incorporated by reference in their entireties. Antibodies can be humanized using a variety of techniques including
20 CDR-grafting (EP 0 239 400; WO 91/09967; US Patent 5,530,101; and 5,585,089), veneering or resurfacing, (EP 0 592 106; EP 0 519 596; Padlan, 1991; Studnicka *et al.*, 1994; Roguska *et al.*, 1994), and chain shuffling (US Patent 5,565,332), which disclosures are hereby incorporated by reference in their entireties. Human antibodies can be made by a variety of methods known in the art including phage display methods described above. *See also*, US Patents 4,444,887, 4,716,111,
25 5,545,806, and 5,814,318; WO 98/46645; WO 98/50433; WO 98/24893; WO 96/34096; WO 96/33735; and WO 91/10741 (said references incorporated by reference in their entireties).

Further included in the present invention are antibodies recombinantly fused or chemically conjugated (including both covalently and non-covalently conjugations) to a polypeptide of the present invention. The antibodies may be specific for antigens other than polypeptides of the
30 present invention. For example, antibodies of the present invention may be recombinantly fused or conjugated to molecules useful as labels in detection assays and effector molecules such as heterologous polypeptides, drugs, or toxins. *See, e.g.*, WO 92/08495; WO 91/14438; WO 89/12624; US Patent 5,314,995; and EP 0 396 387, which disclosures are hereby incorporated by reference in their entireties. Fused antibodies may also be used to target the polypeptides of the present
35 invention to particular cell types, either *in vitro* or *in vivo*, by fusing or conjugating the polypeptides of the present invention to antibodies specific for particular cell surface receptors. Antibodies fused or conjugated to the polypeptides of the present invention may also be used in *vitro* immunoassays and purification methods using methods known in the art (*See e.g.*, Harbor *et al. supra*; WO

93/21232; EP 0 439 095; Naramura, M. *et al.* 1994; US Patent 5,474,981; Gillies *et al.*, 1992; Fell *et al.*, 1991) (said references incorporated by reference in their entireties).

The present invention further includes compositions comprising the polypeptides of the present invention fused or conjugated to antibody domains other than the variable regions. For example, the polypeptides of the present invention may be fused or conjugated to an antibody Fc region, or portion thereof. The antibody portion fused to a polypeptide of the present invention may comprise the hinge region, CH1 domain, CH2 domain, and CH3 domain or any combination of whole domains or portions thereof. The polypeptides of the present invention may be fused or conjugated to the above antibody portions to increase the *in vivo* half-life of the polypeptides or for use in immunoassays using methods known in the art. The polypeptides may also be fused or conjugated to the above antibody portions to form multimers. For example, Fc portions fused to the polypeptides of the present invention can form dimers through disulfide bonding between the Fc portions. Higher multimeric forms can be made by fusing the polypeptides to portions of IgA and IgM. Methods for fusing or conjugating the polypeptides of the present invention to antibody portions are known in the art. *See e.g.*, US Patents 5,336,603, 5,622,929, 5,359,046, 5,349,053, 5,447,851, 5,112,946; EP 0 307 434, EP 0 367 166; WO 96/04388, WO 91/06570; Ashkenazi *et al.* (1991); Zheng *et al.* (1995); and Vil *et al.* (1992) (said references incorporated by reference in their entireties).

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of GENSET than the one to which antibody binding is desired, and animals which do not express GENSET (i.e. a GENSET knock out animal as described herein) are particularly useful for preparing antibodies. GENSET knock out animals will recognize all or most of the exposed regions of a GENSET protein as foreign antigens, and therefore produce antibodies with a wider array of GENSET epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to any one of the GENSET proteins. In addition, the humoral immune system of animals which produce a species of GENSET that resembles the antigenic sequence will preferentially recognize the differences between the animal's native GENSET species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique will be particularly useful in obtaining antibodies that specifically bind to any one of the GENSET proteins.

The antibodies of the invention may be labeled by, e.g., any one of the radioactive, fluorescent or enzymatic labels known in the art.

USES OF POLYNUCLEOTIDES

Uses of polynucleotides as reagents

The polynucleotides of the present invention, particularly those described in the "Oligonucleotide primers and probes" section, may be used as reagents in isolation procedures,

diagnostic assays, and forensic procedures. For example, sequences from the GENSET polynucleotides of the invention may be detectably labeled and used as probes to isolate other sequences capable of hybridizing to them. In addition, sequences from the GENSET polynucleotides of the invention may be used to design PCR primers to be used in isolation, 5 diagnostic, or forensic procedures.

In forensic analyses

PCR primers may be used in forensic analyses, such as the DNA fingerprinting techniques described below. Such analyses may utilize detectable probes or primers based on the sequences of the polynucleotides of the invention. Consequently, the present invention encompasses methods of 10 identification of an individual using the polynucleotides of the invention in forensic analyses, wherein said method includes the steps of:

- a) obtaining a biological sample containing nucleic acid material from an individual;
- b) obtaining an identification pattern for this individual using the polynucleotides of the invention, particularly using GENSET primers and probes;
- 15 c) comparing said identification pattern with a reference identification pattern; and
- d) determining whether said identification pattern is identical to said reference identification pattern.

In one embodiment of this method, the identification pattern consists in sequences of amplicons obtained using GENSET primers as explained in the sections entitled "Forensic 20 Matching by DNA Sequencing" and "Positive Identification by DNA Sequencing".

In another embodiment, the identification pattern consists in unique band or dot patterns obtained using any method described in the sections entitled "Southern Blot Forensic Identification", "Dot Blot Identification Procedure" and "Alternative "Fingerprint" Identification Technique".

25 Table I provides sets of related cDNAs of the invention, e.g. sequences that represent allelic variants of a single sequence. Such variants are especially useful for the herein-described forensic analyses, and are also useful as polymorphic markers to examine, e.g. associations between the herein-discussed GENSET genes and various diseases or conditions.

Forensic Matching by DNA Sequencing

30 In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers designed from different polynucleotides of the invention using any technique known to those skilled in the art including those described herein, is then utilized to amplify DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test 35 subject. Each of these identification DNAs is then sequenced using standard techniques, and a simple database comparison determines the differences, if any, between the sequences from the subject and those from the sample. Statistically significant differences between the suspect's DNA

sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the
5 suspect and the sample.

Positive Identification by DNA Sequencing

The "Forensic Matching by DNA Sequencing" technique described herein may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of polynucleotides of the invention.
10 Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question. Each of these DNA segments is sequenced. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that
15 individual.

Southern Blot Forensic Identification

The "Positive Identification by DNA Sequencing" procedure described herein is repeated to obtain a panel of at least 10 amplified sequences from an individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel contains 100 amplified
20 sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to
25 those with skill in the art. For a review of Southern blotting see Davis *et al.* (1986), which disclosure is hereby incorporated by reference in its entirety.

A panel of probes based on the sequences of the polynucleotides of the invention, or fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using
30 techniques known in the art. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the polynucleotide of the invention. More preferably, the probe comprises at least 20-30 consecutive nucleotides from the polynucleotide of the invention. In some embodiments, the probe comprises more than 30 nucleotides from the polynucleotide of the invention. In other embodiments, the probe comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at
35 least 200 consecutive nucleotides from the polynucleotide of the invention.

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the

hybridization of a large sample of polynucleotide of the invention will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of cDNA probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

Dot Blot Identification Procedure

Another technique for identifying individuals using the polynucleotide sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10, preferably 50 sequences from the polynucleotide of the invention. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P^{32} using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis *et al.* 1986). The ^{32}P labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood *et al.*, 1985). A unique pattern of dots distinguishes one individual from another individual.

Alternative "Fingerprint" Identification Technique

In a representative alternative fingerprinting procedure, the probes are derived from cDNAs. Preferably, a plurality of probes having sequences from different genes are used as follows. Polynucleotides containing at least 10 consecutive bases from these sequences can be used as probes. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the polynucleotide of the invention. More preferably, the probe comprises at least 20-30 consecutive nucleotides from the polynucleotide of the invention. In some embodiments, the probe comprises more than 30 nucleotides from the polynucleotide of the invention. In other embodiments, the probe comprises at least 40, at least 50, at least 75, at least 100, at least 150, or at least 200 consecutive nucleotides from the polynucleotide of the invention.

Oligonucleotides, generally 20-mers, are prepared from a large number, e.g. 50, 100, or 200, of polynucleotides of the invention using commercially available oligonucleotide services such as Genset (Paris, France). Cell samples from the test subject are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate

polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using standard Southern blotting techniques.

10 ng of each of the oligonucleotides are pooled and end-labeled with P³². The
5 nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

10 To find corresponding genomic DNA sequences

The GENSET cDNAs of the invention may also be used to clone sequences located upstream of the cDNAs of the invention on the corresponding genomic DNA. Such upstream sequences may be capable of regulating gene expression, including promoter sequences, enhancer sequences, and other upstream sequences which influence transcription or translation levels. Once
15 identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion.

Use of cDNAs or Fragments thereof to Clone Upstream Sequences from Genomic DNA

Sequences derived from polynucleotides of the inventions may be used to isolate the
20 promoters of the corresponding genes using chromosome walking techniques. In one chromosome walking technique, which utilizes the GenomeWalker™ kit available from Clontech, five complete genomic DNA samples are each digested with a different restriction enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion, oligonucleotide adapters are ligated to each end of the resulting genomic DNA fragments.

25 For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions (which are incorporated herein by reference) using an outer adaptor primer provided in the kit and an outer gene specific primer. The gene specific primer should be selected to be specific for the polynucleotide of the invention of interest and should have a melting temperature, length, and location in the polynucleotide of the invention which is consistent with its
30 use in PCR reactions. Each first PCR reaction contains 5ng of genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer adaptor primer and outer gene specific primer, 1.1 mM of Mg(OAc)₂, and 1 µl of the Tth polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction is as follows: 1 min at 94 degrees Celsius / 2 sec at 94 degree Celsius, 3 min at 72 degrees Celsius (7 cycles) / 2 sec at 94 degrees Celsius, 3 min at 67
35 degrees Celsius (32 cycles) / 5 min at 67 degrees Celsius.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested primers which are

located internally on the amplicon resulting from the first PCR reaction. For example, 5 µl of the reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 µl volume having a composition identical to that of the first PCR reaction except the nested primers are used. The first nested primer is specific for the adaptor, and is provided with the
5 GenomeWalker™ kit. The second nested primer is specific for the particular polynucleotide of the invention for which the promoter is to be cloned and should have a melting temperature, length, and location in the polynucleotide of the invention which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min at 94 degrees Celsius / 2 sec at 94 degrees Celsius, 3 min at 72 degrees Celsius (6 cycles) / 2 sec at 94 degrees Celsius, 3 min
10 at 67 degrees Celsius (25 cycles) / 5 min at 67 degrees Celsius

The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques. Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising
15 at least 15 nucleotides from the polynucleotide of the invention sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the polynucleotide of the invention sequence are isolated as described herein. Thereafter, the single stranded DNA containing the polynucleotide of the invention sequence is released from the beads and converted into double stranded DNA using a primer specific for the
20 polynucleotide of the invention sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the GENSET polynucleotide sequences are identified by colony PCR or colony hybridization.

Identification of Promoters in Cloned Upstream Sequences

Once the upstream genomic sequences have been cloned and sequenced as described above,
25 prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the polynucleotides of the inventions with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as follows. The expression of the reporter gene will be detected when placed under the
30 control of regulatory active polynucleotide fragments or variants of the GENSET promoter region located upstream of the first exon of the GENSET gene. Suitable promoter reporter vectors, into which the GENSET promoter sequences may be cloned include pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech, or pGL2-basic or pGL3-basic promoterless luciferase reporter gene vector from Promega. Briefly,
35 each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, luciferase, beta-galactosidase, or green fluorescent protein. The sequences upstream the GENSET coding region are inserted into the cloning sites upstream of the reporter gene in both orientations and

introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be
5 cloned into vectors which contain an enhancer for increasing transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Promoter sequence within the upstream genomic DNA may be further defined by site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the
10 art. For example, the boundaries of promoters may be further investigated by constructing nested 5' and/or 3' deletions in the upstream DNA using conventional techniques such as Exonuclease III or appropriate restriction endonuclease digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has increased, reduced or illuminated promoter activity, such as described, for example, by Coles *et al.* (1998), the disclosure
15 of which is incorporated herein by reference in its entirety. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into cloning sites in
20 promoter reporter vectors. This type of assay is well known to those skilled in the art and is described in WO 97/17359, US Patent No. 5,374,544; EP 582 796; US Patent No. 5,698,389; US 5,643,746; US Patent No. 5,502,176; and US Patent 5,266,488; the disclosures of which are incorporated by reference herein in their entirety.

The strength and the specificity of the promoter of each GENSET gene can be assessed
25 through the expression levels of a detectable polynucleotide operably linked to the GENSET promoter in different types of cells and tissues. The detectable polynucleotide may be either a polynucleotide that specifically hybridizes with a predefined oligonucleotide probe, or a polynucleotide encoding a detectable protein, including a GENSET polypeptide or a fragment or a variant thereof. This type of assay is well known to those skilled in the art and is described in US
30 Patent No. 5,502,176; and US Patent No. 5,266,488; the disclosures of which are incorporated by reference herein in their entirety. Some of the methods are discussed in more detail elsewhere in the application.

The promoters and other regulatory sequences located upstream of the polynucleotides of the inventions may be used to design expression vectors capable of directing the expression of an
35 inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described herein. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of a polynucleotide of the invention derived from an mRNA which is expressed at a high level in muscle

may be used in the expression vector. Such vectors are described in more detail elsewhere in the application.

Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive
5 expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

10 Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

To find similar sequences

Polynucleotides of the invention may be used to isolate and/or purify nucleic acids similar
15 thereto using any methods well known to those skilled in the art including the techniques based on hybridization or on amplification described in this section. These methods may be used to obtain the genomic DNAs which encode the mRNAs from which the GENSET cDNAs are derived, mRNAs corresponding to GENSET cDNAs, or nucleic acids which are homologous to GENSET cDNAs or fragments thereof, such as variants, species homologues or orthologs. Thus, a plurality of cDNAs
20 similar to GENSET polynucleotides may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or used in diagnostic assays as described herein. cDNAs prepared by any method described therein may be subsequently engineered to obtain nucleic acids which include desired fragments of the cDNA using conventional techniques such as subcloning, PCR, or *in vitro* oligonucleotide synthesis. For example, nucleic acids which include only the coding sequences
25 may be obtained using techniques known to those skilled in the art. Similarly, nucleic acids containing any other desired fragment of the coding sequences for the encoded protein may be obtained.

Indeed, cDNAs of the present invention or fragments thereof may be used to isolate nucleic acids similar to cDNAs from a cDNA library or a genomic DNA library. Such cDNA libraries or
30 genomic DNA libraries may be obtained from a commercial source or made using techniques familiar to those skilled in the art such as those described in PCT publication WO 00/37491, which disclosure is hereby incorporated by reference in its entirety. Examples of methods for obtaining nucleic acids similar to GENSET polynucleotides are described below.

Hybridization-based methods

35 Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook *et al.*, (1989) and in Hames and Higgins (1985), the

disclosures of which are incorporated herein by reference in their entireties. The same techniques may be used to isolate genomic DNAs.

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. Any polynucleotide fragment of the invention may be used as a probe, in particular those defined in the "Oligonucleotide primers and probes" section. A probe comprising at least 10 consecutive nucleotides from a GENSET cDNA or fragment thereof is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the cDNA or fragment thereof. More preferably, the probe comprises 20 to 30 consecutive nucleotides from the cDNA or fragment thereof. In some embodiments, the probe comprises more than 30 nucleotides from the cDNA or fragment thereof.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

By varying the stringency of the hybridization conditions used to identify cDNAs or genomic DNAs which hybridize to the detectable probe, cDNAs or genomic DNAs having different levels of identity to the probe can be identified and isolated as described below.

Stringent conditions

"Stringent hybridization conditions" are defined as conditions in which only nucleic acids having a high level of identity to the probe are able to hybridize to said probe. These conditions may be calculated as follows:

For probes between 14 and 70 nucleotides in length the melting temperature (T_m) is calculated using the formula: $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (600/N)$ where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation: $T_m = 81.5 + 16.6(\log(Na^+)) + 0.41(\text{fraction G+C}) - (0.63\% \text{ formamide}) - (600/N)$ where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μ g denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μ g denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, 1986.

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient

period of time to allow the probe to hybridize to nucleic acids containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the T_m. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the T_m. Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

Nucleic acids which have hybridized to the probe are identified by autoradiography or other conventional techniques.

Low and moderate conditions

Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. The above procedure may thus be modified to identify nucleic acids having decreasing levels of identity to the probe sequence. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C. Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of identity to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide. cDNAs or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Consequently, the present invention encompasses methods of isolating nucleic acids similar to the polynucleotides of the invention, comprising the steps of:

a) contacting a collection of cDNA or genomic DNA molecules with a detectable probe comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40 or 50 consecutive nucleotides of a sequence

selected from the group consisting of the sequences of SEQ ID NOs: 1-169, 339-455, 561-784, the sequences of clones inserts of the deposited clone pool and sequences complementary thereto under stringent, moderate or low conditions which permit said probe to hybridize to at least a cDNA or genomic DNA molecule in said collection;

- 5 b) identifying said cDNA or genomic DNA molecule which hybridizes to said detectable probe; and
- c) isolating said cDNA or genomic DNA molecule which hybridized to said probe.

PCR-based methods

In addition to the above described methods, other protocols are available to obtain
10 homologous cDNAs using GENSET cDNA of the present invention or fragment thereof as outlined in the following paragraphs.

cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing polyA selection procedures or other techniques known to those skilled in the art. A first primer capable of hybridizing to the polyA tail of the
15 mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a first cDNA strand.

The term "capable of hybridizing to the polyA tail of said mRNA" refers to and embraces all primers containing stretches of thymidine residues, so-called oligo(dT) primers, that hybridize to the 3' end of eukaryotic poly(A)+ mRNAs to prime the synthesis of a first cDNA strand.
20 Techniques for generating said oligo (dT) primers and hybridizing them to mRNA to subsequently prime the reverse transcription of said hybridized mRNA to generate a first cDNA strand are well known to those skilled in the art and are described in Current Protocols in Molecular Biology, John Wiley and Sons, Inc. 1997 and Sambrook *et al.*, 1989. Preferably, said oligo (dT) primers are present in a large excess in order to allow the hybridization of all mRNA 3'ends to at least one oligo
25 (dT) molecule. The priming and reverse transcription steps are preferably performed between 37°C and 55°C depending on the type of reverse transcriptase used. Preferred oligo(dT) primers for priming reverse transcription of mRNAs are oligonucleotides containing a stretch of thymidine residues of sufficient length to hybridize specifically to the polyA tail of mRNAs, preferably of 12 to 18 thymidine residues in length. More preferably, such oligo(T) primers comprise an additional
30 sequence upstream of the poly(dT) stretch in order to allow the addition of a given sequence to the 5'end of all first cDNA strands which may then be used to facilitate subsequent manipulation of the cDNA. Preferably, this added sequence is 8 to 60 residues in length. For instance, the addition of a restriction site in 5' of cDNAs facilitates subcloning of the obtained cDNA. Alternatively, such an added 5' end may also be used to design primers of PCR to specifically amplify cDNA clones of
35 interest.

The first cDNA strand is then hybridized to a second primer. Any polynucleotide fragment of the invention may be used, and in particular those described in the "Oligonucleotide primers and probes" section. This second primer contains at least 10 consecutive nucleotides of a polynucleotide

of the invention. Preferably, the primer comprises at least 10, 12, 15, 17, 18, 20, 23, 25, or 28 consecutive nucleotides of a polynucleotide of the invention. In some embodiments, the primer comprises more than 30 nucleotides of a polynucleotide of the invention. If it is desired to obtain cDNAs containing the full protein coding sequence, including the authentic translation initiation site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RT-PCR may be performed as described above using primers from both ends of the cDNA to be obtained.

The double stranded cDNAs made using the methods described above are isolated and cloned. The cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to generate a first cDNA strand, extending a primer to make a second cDNA strand complementary to the first cDNA strand, isolating the double stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described in *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. 1997 and Sambrook *et al.*, 1989.

Consequently, the present invention encompasses methods of making cDNAs. In a first embodiment, the method of making a cDNA comprises the steps of

- a) contacting a collection of mRNA molecules from human cells with a primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of the sequences complementary to SEQ ID NOS:1-169, 339-455, 561-784 and sequences complementary to a clone insert of the deposited clone pool;
- b) hybridizing said primer to an mRNA in said collection;
- c) reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA;
- d) making a second cDNA strand complementary to said first cDNA strand; and
- e) isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

Another embodiment of the present invention is a purified cDNA obtainable by the method of the preceding paragraph. In one aspect of this embodiment, the cDNA encodes at least a portion of a human polypeptide.

In a second embodiment, the method of making a cDNA comprises the steps of

- a) contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA;
- b) hybridizing said first primer to said polyA tail;
- c) reverse transcribing said mRNA to make a first cDNA strand;
- d) making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of

a sequence selected from the group consisting of SEQ ID NOs: 1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool; and

e) isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

5 In another aspect of this method the second cDNA strand is made by

a) contacting said first cDNA strand with a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool, and a third primer which sequence is fully included within the sequence of said first primer;

10 b) performing a first polymerase chain reaction with said second and third primers to generate a first PCR product;

c) contacting said first PCR product with a fourth primer, comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of said sequence selected from the group consisting of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited
15 clone pool, and a fifth primer, which sequence is fully included within the sequence of said third primer, wherein said fourth and fifth hybridize to sequences within said first PCR product; and

d) performing a second polymerase chain reaction, thereby generating a second PCR product.

Alternatively, the second cDNA strand may be made by contacting said first cDNA strand
20 with a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool, and a third primer which sequence is fully included within the sequence of said first primer and performing a polymerase chain reaction with said second and third primers to generate said second cDNA strand.

25 Alternatively, the second cDNA strand may be made by:

a) contacting said first cDNA strand with a second primer comprising at least 12, 15, 18, 20, 23, 25, 28, 30, 35, 40, or 50 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone
pool;

30 b) hybridizing said second primer to said first strand cDNA; and

c) extending said hybridized second primer to generate said second cDNA strand.

Another embodiment of the present invention is a purified cDNA obtainable by a method of making a cDNA of the invention. In one aspect of this embodiment, said cDNA encodes at least a portion of a human polypeptide.

35 *Other protocols*

Alternatively, other procedures may be used for obtaining homologous cDNAs. In one approach, cDNAs are prepared from mRNA and cloned into double stranded phagemids as follows. The cDNA library in the double stranded phagemids is then rendered single stranded by treatment

with an endonuclease, such as the Gene II product of the phage F1 and an exonuclease (Chang *et al.*, 1993, which disclosure is hereby incorporated by reference in its entirety). A biotinylated oligonucleotide comprising the sequence of a fragment of a known GENSET cDNA, genomic DNA or fragment thereof is hybridized to the single stranded phagemids. Preferably, the fragment
5 comprises at least 10, 12, 15, 17, 18, 20, 23, 25, or 28 consecutive nucleotides of a sequence selected from the group consisting of the sequences of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited clone pool.

Hybrids between the biotinylated oligonucleotide and phagemids are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet
10 (Fry *et al.*, 1992, which disclosure is hereby incorporated by reference in its entirety). Thereafter, the resulting phagemids are released from the beads and converted into double stranded DNA using a primer specific for the GENSET cDNA or fragment used to design the biotinylated oligonucleotide. Alternatively, protocols such as the Gene Trapper kit (Gibco BRL), which disclosure is which disclosure is hereby incorporated by reference in its entirety, may be used. The
15 resulting double stranded DNA is transformed into bacteria. Homologous cDNAs to the GENSET cDNA or fragment thereof sequence are identified by colony PCR or colony hybridization.

As a chromosome marker

Chromosomal localization of the cDNA of the present invention were determined using information from public and proprietary databases. Table II lists the putative chromosomal location
20 of the polynucleotides of the present invention. Column one lists the sequence identification number with the corresponding chromosomal location listed in column two. Thus, the present invention also relates to methods and compositions using the chromosomal location of the polynucleotides of the invention to construct a human high resolution map or to identify a given chromosome in a sample using any techniques known to those skilled in the art including those disclosed below.

25 GENSET polynucleotides may also be mapped to their chromosomal locations using any methods or techniques known to those skilled in the art including radiation hybrid (RH) mapping, PCR-based mapping and Fluorescence in situ hybridization (FISH) mapping described below.

Radiation hybrid mapping

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high
30 resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones containing different fragments of the human genome. This technique is described by Benham *et al.* (1989) and Cox *et al.*, (1990), which disclosures are hereby
35 incorporated by reference in their entireties. The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering GENSET cDNAs or genomic DNAs. In this

approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler *et al.*, 1996), which disclosure is hereby incorporated by reference in its entirety.

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK) (Foster *et al.*, 1996), the region surrounding the Gorlin syndrome gene (Obermayr *et al.*, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers *et al.*, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer *et al.*, 1992) and 13 loci on the long arm of chromosome 5 (Warrington *et al.*, 1991), which disclosures are hereby incorporated by reference in their entireties.

Mapping of cDNAs to Human Chromosomes using PCR techniques

GENSET cDNAs and genomic DNAs may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the cDNA sequence to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich (1992), which disclosure is hereby incorporated by reference in its entirety.

The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 uCi of a ³²P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94 degrees Celsius, 1.4 min; 55 degrees Celsius, 2 min; and 72 degrees Celsius, 2 min; with a final extension at 72 degrees Celsius for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the cDNA from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given cDNA or genomic DNA. DNA is isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from the GENSET cDNAs or genomic DNAs. Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the GENSET cDNA or genomic DNA will yield an amplified fragment. The GENSET cDNAs or genomic DNAs are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the

chromosome containing that GENSET cDNA or genomic DNA. For a review of techniques and analysis of results from somatic cell gene mapping experiments, see Ledbetter *et al.*, (1990), which disclosure is hereby incorporated by reference in its entirety.

Mapping of cDNAs to Chromosomes Using Fluorescence in situ Hybridization

5 Fluorescence *in situ* hybridization (FISH) allows the GENSET cDNA or genomic DNA to be mapped to a particular location on a given chromosome. The chromosomes to be used for fluorescence *in situ* hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

In a preferred embodiment, chromosomal localization of a GENSET cDNA or genomic
10 DNA is obtained by FISH as described by Cherif *et al.* (1990), which disclosure is hereby incorporated by reference in its entirety. Metaphase chromosomes are prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10 uM) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BudR, 0.1 mM) for 6 h.
15 Colcemid (1 ug/ml) is added for the last 15 min before harvesting the cells. Cells are collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37 degrees Celsius for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The GENSET cDNA or genomic DNA is labeled with biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research Laboratories,
20 Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upssala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70 degrees Celsius for 5-10 min.

Slides kept at -20 degrees Celsius are treated for 1 h at 37 degrees Celsius with RNase A
25 (100 ug/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70 degrees Celsius, then dehydrated at 4 degrees Celsius. The slides are treated with proteinase K (10 ug/100 ml in 20 mM Tris-HCl, 2 mM CaCl₂) at 37 degrees Celsius for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a coverslip, sealed with rubber cement and
30 incubated overnight in a humid chamber at 37 degrees Celsius. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of biotinylated goat anti-avidin and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif *et al.*, 1990). The slides are observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained
35 with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular GENSET cDNA or genomic DNA may be localized to a particular cytogenetic R-band on a given chromosome.

Use of cDNAs to Construct or Expand Chromosome Maps

Once the GENSET cDNAs or genomic DNAs have been assigned to particular chromosomes using any technique known to those skilled in the art those skilled in the art, particularly those described herein, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome.

- One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the GENSET cDNAs or genomic DNAs are obtained. This approach is described in Nagaraja *et al.* (1997), which disclosure is hereby incorporated by reference in its entirety. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they include the GENSET cDNA or genomic DNA whose position is to be determined. Once an insert has been found which includes the GENSET cDNA or genomic DNA, the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the GENSET cDNA or genomic DNA was derived. This process can be repeated for each insert in the YAC library to determine the location of each of the GENSET cDNA or genomic DNA relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

Identification of genes associated with hereditary diseases or drug response

- This example illustrates an approach useful for the association of GENSET cDNAs or genomic DNAs with particular phenotypic characteristics. In this example, a particular GENSET cDNA or genomic DNA is used as a test probe to associate that GENSET cDNA or genomic DNA with a particular phenotypic characteristic.

- GENSET cDNAs or genomic DNAs are mapped to a particular location on a human chromosome using techniques such as those described herein or other techniques known in the art. A search of Mendelian Inheritance in Man (V. McKusick, Mendelian Inheritance in Man; available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human chromosome which contains the GENSET cDNA or genomic DNA to be a very gene rich region containing several known genes and several diseases or phenotypes for which genes have not been identified. The gene corresponding to this GENSET cDNA or genomic DNA thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the GENSET cDNA or genomic DNA are used to screen genomic DNA, mRNA or cDNA obtained from the patients. GENSET cDNAs or genomic DNAs that are not amplified in the patients can be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when the samples are derived from an individual having the phenotype associated with the disease than when the sample is derived from a healthy individual, indicating that the gene containing the cDNA may be responsible for the genetic disease.

Uses of polynucleotides in recombinant vectors

The present invention also relates to recombinant vectors including the isolated polynucleotides of the present invention, and to host cells recombinant for a polynucleotide of the invention, such as the above vectors, as well as to methods of making such vectors and host cells and for using them for production of GENSET polypeptides by recombinant techniques.

Recombinant Vectors

The term “vector” is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism. The present invention encompasses a family of recombinant vectors that comprise a regulatory polynucleotide and/or a coding polynucleotide derived from either the GENSET genomic sequence or the cDNA sequence. Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences, coding sequences and polynucleotide constructs, as well as any GENSET primer or probe as defined herein.

In a first preferred embodiment, a recombinant vector of the invention is used to amplify the inserted polynucleotide derived from a GENSET genomic sequence or a GENSET cDNA, for example any cDNA selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784, sequences of clone inserts of the deposited clone pool, variants and fragments thereof in a suitable cell host, this polynucleotide being amplified at every time that the recombinant vector replicates.

A second preferred embodiment of the recombinant vectors according to the invention comprises expression vectors comprising either a regulatory polynucleotide or a coding nucleic acid of the invention, or both. Within certain embodiments, expression vectors are employed to express a GENSET polypeptide which can be then purified and, for example be used in ligand screening assays or as an immunogen in order to raise specific antibodies directed against the GENSET protein. In other embodiments, the expression vectors are used for constructing transgenic animals and also for gene therapy. Expression requires that appropriate signals are provided in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and

mammalian sources that drive expression of the genes of interest in host cells. Dominant drug selection markers for establishing permanent, stable cell clones expressing the products are generally included in the expression vectors of the invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

5 More particularly, the present invention relates to expression vectors which include nucleic acids encoding a GENSET protein, preferably a GENSET protein with an amino acid sequence selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918, sequences of polypeptides encoded by the clone inserts of the deposited clone pool, variants and fragments thereof. The polynucleotides of the present invention may be used to express an encoded
10 protein in a host organism to produce a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the host organism or stably expressed in the host organism. The encoded protein may have any of the activities described herein. The encoded protein may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

15 Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

General features of the expression vectors of the invention

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a
20 cosmid, a plasmid or even a linear DNA molecule which may comprise a chromosomal, non-chromosomal, semi-synthetic and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of:

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300
25 bp in length that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, said structural or coding sequence being operably linked to the regulatory elements described in (1); and

(3) appropriate transcription initiation and termination sequences. Structural units intended
30 for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, when a recombinant protein is expressed without a leader or transport sequence, it may include a N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

35 Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences,

and preferably a leader sequence capable of directing secretion of the translated protein into the periplasmic space or the extracellular medium. In a specific embodiment wherein the vector is adapted for transfecting and expressing desired sequences in mammalian host cells, preferred vectors will comprise an origin of replication in the desired host, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation signals, splice donor and acceptor sites, transcriptional termination sequences, and 5'-flanking non-transcribed sequences. DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter, enhancer, splice and polyadenylation signals may be used to provide the required non-transcribed genetic elements.

10 The *in vivo* expression of a GENSET polypeptide of the present invention may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism, for the treatment or prevention of any disease or condition that can be treated or prevented by increasing the level of GENSET polypeptide expression, or to the production of a biologically inactive GENSET protein. Consequently, the present invention also comprises recombinant
15 expression vectors mainly designed for the *in vivo* production of a GENSET polypeptide the present invention by the introduction of the appropriate genetic material in the organism or the patient to be treated. This genetic material may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue.

20 *Regulatory Elements*

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account the cell host in which the heterologous gene has to be expressed. The particular promoter employed to control the expression of a nucleic acid sequence of interest is not believed to be important, so long as it is capable of directing the expression of the
25 nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell, such as, for example, a human or a viral promoter.

A suitable promoter may be heterologous with respect to the nucleic acid for which it controls the expression or alternatively can be endogenous to the native polynucleotide containing
30 the coding sequence to be expressed. Additionally, the promoter is generally heterologous with respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors.

35 Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the gpt, lambda PR, PL and trp promoters (EP 0036776), the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen), (Smith *et al.*, 1983;

O'Reilly *et al.*, 1992; which disclosures are hereby incorporated by reference in their entireties), the lambda PR promoter or also the trc promoter.

Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and
5 promoter is well within the level of ordinary skill in the art. The choice of a promoter is well within the ability of a person skilled in the field of genetic engineering. For example, one may refer to the book of Sambrook *et al.*, (1989) or also to the procedures described by Fuller *et al.*, (1996), which disclosures are hereby incorporated by reference in their entireties.

Other regulatory elements

10 Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to
15 enhance message levels and to minimize read through from the cassette into other sequences.

Selectable Markers

Selectable markers confer an identifiable change to the cell permitting easy identification of cells containing the expression construct. The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture,
20 TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. Coli*, or levan saccharase for mycobacteria, this latter marker being a negative selection marker.

Preferred Vectors

Bacterial vectors

As a representative but non-limiting example, useful expression vectors for bacterial use
25 can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and pGEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of other suitable vectors are known to those of skill in the art, and
30 commercially available, such as the following bacterial vectors: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

Bacteriophage vectors

The P1 bacteriophage vector may contain large inserts ranging from about 80 to about 100 kb. The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg (1992, 1994), which disclosure is hereby incorporated by reference in its entirety.

- 5 Recombinant P1 clones comprising GENSET nucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (See Linton *et al.*, 1993), which disclosure is hereby incorporated by reference in its entirety. To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick *et al.* (1994), which disclosure is hereby incorporated by reference in its entirety. Briefly, *E. coli* (preferably strain NS3529) harboring the
- 10 P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating
- 15 the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

- When the goal is to express a P1 clone comprising GENSET polypeptide-encoding nucleotide sequences in a transgenic animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting
- 20 sites within the P1 polylinker (*Sfi*I, *Not*I or *Sal*I). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar to those originally reported for the isolation of DNA from YACs (See e. g., Schedl *et al.*, 1993a; Peterson *et al.*, 1993), which disclosures are hereby incorporated by reference in their entireties. At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore Ultrafree-MC Filter Unit
- 25 (Millipore, Bedford, MA, USA – 30,000 molecular weight limit) and then dialyzed against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250 µM EDTA) containing 100 mM NaCl, 30 µM spermine, 70 µM spermidine on a microdialysis membrane (type VS, 0.025 µM from Millipore). The intactness of the purified P1 DNA insert is assessed by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

30 Viral vectors

- In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus vectors according to the invention are those described by Feldman and Steg (1996), or Ohno *et al.*, (1994), which disclosures are hereby incorporated by reference in their entireties. Another preferred recombinant adenovirus according to this specific embodiment of the present invention is
- 35 the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent application No. FR-93.05954), which disclosure is hereby incorporated by reference in its entirety.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery systems of choice for the transfer of exogenous polynucleotides *in vivo*,

particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected from the group

5 consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include the 4070A and the 1504A viruses, Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include

10 Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728). Other preferred retroviral vectors are those described in Roth *et al.* (1996), PCT Application No WO 93/25234, PCT Application No WO 94/ 06920, Roux *et al.*, (1989), Julian *et al.*, (1992), and Neda *et al.*, (1991), which disclosures are hereby incorporated by reference in their entireties.

Yet another viral vector system that is contemplated by the invention comprises the adeno-

15 associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka *et al.*, 1992), which disclosure is hereby incorporated by reference in its entirety. It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte *et al.* 1992;

20 Samulski *et al.*, 1989; McLaughlin *et al.*, 1989), which disclosures are hereby incorporated by reference in their entireties. One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

BAC vectors

The bacterial artificial chromosome (BAC) cloning system (Shizuya *et al.*, 1992), which

25 disclosure is hereby incorporated by reference in its entirety, has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in *E. coli*. A preferred BAC vector comprises a pBeloBAC11 vector that has been described by Kim *et al.* (1996), which disclosure is hereby incorporated by reference in its entirety. BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into

30 either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The

35 cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be linearized by treatment of the BAC vector with the commercially available enzyme lambda

terminase that leads to the cleavage at the unique *cosN* site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

Baculovirus

Another specific suitable host vector system is the pVL1392/1393 baculovirus transfer
5 vector (Pharming) that is used to transfect the SF9 cell line (ATCC No. CRL 1711) which is derived from *Spodoptera frugiperda*. Other suitable vectors for the expression of the GENSET polypeptide of the present invention in a baculovirus expression system include those described by Chai *et al.*, (1993), Vlasak *et al.*, (1983), and Lenhard *et al.*, (1996), which disclosures are hereby incorporated by reference in their entireties.

10 Delivery Of The Recombinant Vectors

To effect expression of the polynucleotides and polynucleotide constructs of the invention, the constructs must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cell lines, or *in vivo* or *ex vivo*, as in the treatment of certain diseases states. One mechanism is viral infection where the expression construct is encapsulated in
15 an infectious viral particle.

Several non-viral methods for the transfer of polynucleotides into cultured mammalian cells are also contemplated by the present invention, and include, without being limited to, calcium phosphate precipitation (Graham *et al.*, 1973; Chen *et al.*, 1987); DEAE-dextran (Gopal, 1985); electroporation (Tur-Kaspa *et al.*, 1986; Potter *et al.*, 1984); direct microinjection (Harland *et al.*,
20 1985); DNA-loaded liposomes (Nicolau *et al.*, 1982; Fraley *et al.*, 1979); and receptor-mediated transfection. (Wu and Wu, 1987, 1988), which disclosures are hereby incorporated by reference in their entireties. Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

Once the expression polynucleotide has been delivered into the cell, it may be stably integrated into the genome of the recipient cell. This integration may be in the cognate location and
25 orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

30 One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect. This is
35 particularly applicable for transfer *in vitro* but it may be applied to *in vivo* as well.

Compositions for use *in vitro* and *in vivo* comprising a "naked" polynucleotide are described in PCT application No. WO 90/11092 (Vical Inc.) and also in PCT application No. WO

95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tascon *et al.* (1996) and of Huygen *et al.*, (1996), which disclosures are hereby incorporated by reference in their entireties.

In still another embodiment of the invention, the transfer of a naked polynucleotide of the invention, including a polynucleotide construct of the invention, into cells may be accomplished with particle bombardment (biolistic), said particles being DNA-coated microprojectiles accelerated to a high velocity allowing them to pierce cell membranes and enter cells without killing them, such as described by Klein *et al.*, (1987), which disclosure is hereby incorporated by reference in its entirety.

In a further embodiment, the polynucleotide of the invention may be entrapped in a liposome (Ghosh and Bacchawat, 1991; Wong *et al.*, 1980; Nicolau *et al.*, 1987, which disclosures are hereby incorporated by reference in their entireties).

In a specific embodiment, the invention provides a composition for the *in vivo* production of the GENSET polypeptides described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0.1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced *in vitro* in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired GENSET polypeptide or the desired fragment thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

Secretion vectors

Some of the GENSET cDNAs or genomic DNAs of the invention may also be used to construct secretion vectors capable of directing the secretion of the proteins encoded by genes inserted in the vectors. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by reducing the number of background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described below.

The secretion vectors of the present invention include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

A signal sequence from a polynucleotide of the invention, preferably a signal sequences selected from the group of signal sequences of SEQ ID NOs: 1-85, 339-400, 406-407, 413-415, 561-594, and 634-651 and signal sequences of clone inserts of the deposited clone pool is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the GENSET cDNA or genomic DNA. Suitable hosts include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Preferably, the secretion vector is maintained in multiple copies in each host cell. As used herein, multiple copies means at least 2, 5, 10, 20, 25, 50 or more than 50 copies per cell. In some embodiments, the multiple copies are maintained extrachromosomally. In other embodiments, the multiple copies result from amplification of a chromosomal sequence.

Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange chromatography, and hplc. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence

such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic effect.

5

Cell Hosts

Another object of the invention comprises a host cell that has been transformed or transfected with one of the polynucleotides described herein, and in particular a polynucleotide either comprising a GENSET polypeptide-encoding polynucleotide regulatory sequence or the
10 polynucleotide coding for a GENSET polypeptide. Also included are host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above. However, the cell hosts of the present invention can comprise any of the polynucleotides of the present invention. In a preferred embodiment, host cells contain a polynucleotide sequence comprising a sequence selected from the group consisting of sequences of
15 SEQ ID NOs:1-169, 339-455, 561-784, sequences of clone inserts of the deposited clone pool, variants and fragments thereof. Preferred host cells used as recipients for the expression vectors of the invention are the following:

a) Prokaryotic host cells: *Escherichia coli* strains (I.E.DH5- α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and
20 *Staphylococcus*.

b) Eukaryotic host cells: HeLa cells (ATCC No.CCL2; No.CCL2.1; No.CCL2.2), Cv 1 cells (ATCC No.CCL70), COS cells (ATCC No.CRL1650; No.CRL1651), Sf-9 cells (ATCC No.CRL1711), C127 cells (ATCC No. CRL-1804), 3T3 (ATCC No. CRL-6361), CHO (ATCC No. CCL-61), human kidney 293. (ATCC No. 45504; No. CRL-1573) and BHK (ECACC No.
25 84100501; No. 84111301).

c) Other mammalian host cells.

The present invention also encompasses primary, secondary, and immortalized homologously recombinant host cells of vertebrate origin, preferably mammalian origin and
30 particularly human origin, that have been engineered to: a) insert exogenous (heterologous) polynucleotides into the endogenous chromosomal DNA of a targeted gene, b) delete endogenous chromosomal DNA, and/or c) replace endogenous chromosomal DNA with exogenous polynucleotides. Insertions, deletions, and/or replacements of polynucleotide sequences may be to the coding sequences of the targeted gene and/or to regulatory regions, such as promoter and
35 enhancer sequences, operably associated with the targeted gene.

In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin, particularly mammalian origin, that have been engineered to delete or replace endogenous genetic

material (e.g., coding sequence), and/or to include genetic material (e.g., heterologous polynucleotide sequences) that is operably associated with the polynucleotides of the invention, and which activates, alters, and/or amplifies endogenous polynucleotides. For example, techniques known in the art may be used to operably associate heterologous control regions (e.g., promoter and/or enhancer) and endogenous polynucleotide sequences via homologous recombination, see, 5 e.g., U.S. Patent No. 5,641,670, issued June 24, 1997; International Publication No. WO 96/29411, published September 26, 1996; International Publication No. WO 94/12650, published August 4, 1994; Koller *et al.*, (1989); and Zijlstra *et al.* (1989) (the disclosures of each of which are incorporated by reference in their entireties).

10 The present invention further relates to a method of making a homologously recombinant host cell *in vitro* or *in vivo*, wherein the expression of a targeted gene not normally expressed in the cell is altered. Preferably the alteration causes expression of the targeted gene under normal growth conditions or under conditions suitable for producing the polypeptide encoded by the targeted gene. The method comprises the steps of: (a) transfecting the cell *in vitro* or *in vivo* with a polynucleotide 15 construct, said polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell *in vitro* or *in vivo* under conditions appropriate for homologous recombination.

The present invention further relates to a method of altering the expression of a targeted 20 gene in a cell *in vitro* or *in vivo* wherein the gene is not normally expressed in the cell, comprising the steps of: (a) transfecting the cell *in vitro* or *in vivo* with a polynucleotide construct, said polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell *in vitro* or *in vivo* under conditions 25 appropriate for homologous recombination, thereby producing a homologously recombinant cell; and (c) maintaining the homologously recombinant cell *in vitro* or *in vivo* under conditions appropriate for expression of the gene.

The present invention further relates to a method of making a polypeptide of the present invention by altering the expression of a targeted endogenous gene in a cell *in vitro* or *in vivo* 30 wherein the gene is not normally expressed in the cell, comprising the steps of: a) transfecting the cell *in vitro* with a polynucleotide construct, said polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; (b) maintaining the transfected cell *in vitro* or *in vivo* under conditions appropriate for homologous recombination, thereby producing a 35 homologously recombinant cell; and c) maintaining the homologously recombinant cell *in vitro* or *in vivo* under conditions appropriate for expression of the gene thereby making the polypeptide.

The present invention further relates to a polynucleotide construct which alters the expression of a targeted gene in a cell type in which the gene is not normally expressed. This occurs when the polynucleotide construct is inserted into the chromosomal DNA of the target cell,

wherein said polynucleotide construct comprises: a) a targeting sequence; b) a regulatory sequence and/or coding sequence; and c) an unpaired splice-donor site, if necessary. Further included are a polynucleotide construct, as described above, wherein said polynucleotide construct further comprises a polynucleotide which encodes a polypeptide and is in-frame with the targeted
5 endogenous gene after homologous recombination with chromosomal DNA.

The compositions may be produced, and methods performed, by techniques known in the art, such as those described in U.S. Patent NOs: 6,054,288; 6,048,729; 6,048,724; 6,048,524; 5,994,127; 5,968,502; 5,965,125; 5,869,239; 5,817,789; 5,783,385; 5,733,761; 5,641,670; 5,580,734 ; International Publication NOs: WO96/29411, WO 94/12650; and scientific articles
10 described by Koller *et al.*, (1994). (The disclosures of each of which are incorporated by reference in their entireties).

GENSET gene expression in mammalian cells, preferably human cells, may be rendered defective, or alternatively may be altered by replacing endogenous GENSET polypeptide-encoding genes in the genome of an animal cell by a GENSET polypeptide-encoding polynucleotide
15 according to the invention. These genetic alterations may be generated by homologous recombination using previously described specific polynucleotide constructs.

Mammal zygotes, such as murine zygotes may be used as cell hosts. For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration ranging from 1 ng/ml –
20 for BAC inserts- to 3 ng/μl –for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7.4, 250 μM EDTA containing 100 mM NaCl, 30 μM spermine, and 70 μM spermidine. When the DNA to be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl *et al* (1993b), which disclosure is hereby incorporated by reference in its entirety.

Any one of the polynucleotides of the invention, including the polynucleotide constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC No.CRL-1821), ES-D3 (ATCC No.CRL1934 and No. CRL-11632), YS001 (ATCC No. CRL-11776), 36.5
30 (ATCC No. CRL-11116). ES cells are maintained in an uncommitted state by culture in the presence of growth-inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells are primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo *et al.*
35 (1993) and are growth-inhibited by irradiation, such as described by Robertson (1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (1990), which disclosures are hereby incorporated by reference in their entireties.

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skilled artisan.

Transgenic Animals

The terms “transgenic animals” or “host animals” are used herein to designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention. In one embodiment, the invention encompasses non-human host mammals and animals comprising a recombinant vector of the invention or a GENSET gene disrupted by homologous recombination with a knock out vector.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a GENSET polypeptide coding sequence, a GENSET polynucleotide regulatory sequence, a polynucleotide construct, or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Generally, a transgenic animal according the present invention comprises any of the polynucleotides, the recombinant vectors and the cell hosts described in the present invention. In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to the dysregulation of the expression of a given GENSET gene, in particular the transgenic animals containing within their genome one or several copies of an inserted polynucleotide encoding a native GENSET polypeptide, or alternatively a mutant GENSET polypeptide.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the GENSET gene, leading to high yields in the synthesis of this protein of interest, and eventually to tissue specific expression of the protein of interest.

The design of the transgenic animals of the invention may be made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to US Patents Nos 4,873,191, issued Oct. 10, 1989; 5,464,764 issued Nov 7, 1995; and 5,789,215, issued Aug 4, 1998; these documents being herein incorporated by reference to disclose methods producing transgenic mice.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material which encodes either a GENSET polypeptide coding sequence, a GENSET polynucleotide regulatory sequence, or a DNA sequence encoding a
5 GENSET polynucleotide antisense sequence, or a portion thereof, such as described in the present specification. A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas *et al.* (1987), which disclosure is hereby incorporated by reference in its entirety. The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by
10 Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour *et al.* (1988), which disclosure is hereby incorporated by reference in its entirety.

15 The positive cells are then isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley (1987), which disclosure is hereby incorporated by reference in its entirety. The blastocysts are then inserted into a female host animal and allowed to grow to term. Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood *et al.* (1993), or by Nagy *et al.* (1993), which
20 disclosures are hereby incorporated by reference in their entireties, the ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line.

The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which ones are wild type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a
25 recombinant expression vector or a recombinant host cell according to the invention.

In another embodiment, transgenic animals are produced by microinjecting polynucleotides
30 are microinjected into a fertilized oocyte. Typically, fertilized oocytes are microinjected using standard techniques, and then cultured in vitro until a "pre-implantation embryo" is obtained. Such pre-implantation embryos preferably contain approximately 16 to 150 cells. Methods for culturing fertilized oocytes to the pre-implantation stage are described, e.g., by Gordon *et al.* ((1984) *Methods in Enzymology*, 101, 414); Hogan *et al.* ((1986) in *Manipulating the mouse embryo. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y) (for the mouse embryo); Hammer *et al.* ((1985) *Nature*, 315, 680) (for rabbit and porcine embryos); Gandolfi *et al.* ((1987) *J. Reprod. Fert.* 81, 23-28); Rexroad *et al.* ((1988) *J. Anim. Sci.* 66, 947-953) (for ovine embryos); and
35 Eyestone *et al.* ((1989) *J. Reprod. Fert.* 85, 715-720); Camous *et al.* ((1984) *J. Reprod. Fert.* 72, 779-785); and Heyman *et al.* ((1987) *Theriogenology* 27, 5968) (for bovine embryos); the disclosures of each of which are incorporated herein in their entireties. Pre-implantation embryos are then transferred to an appropriate female by standard methods to permit the birth of a transgenic or chimeric animal, depending upon the stage of development when the transgene is introduced.

As the frequency of transgene incorporation is often low, the detection of transgene integration in pre-implantation embryos is often desirable using any of the herein-described methods. Any of a number of methods can be used to detect the presence of a transgene in a pre-implantation embryo. For example, one or more cells may be removed from the pre-implantation embryo, and the presence or absence of the transgene in the removed cell or cells can be detected using any standard method e.g. PCR. Alternatively, the presence of a transgene can be detected in utero or post partum using standard methods.

In a particularly preferred embodiment of the present invention, transgenic mammals are generated that secrete recombinant GENSET polypeptides in their milk. As the mammary gland is a highly efficient protein-producing organ, such methods can be used to produce protein concentrations in the gram per liter range, and often significantly more. Preferably, expression in the mammary gland is accomplished by operably linking the polynucleotide encoding the GENSET polypeptide to a mammary gland specific promoter and, optionally, other regulatory elements. Suitable promoters and other elements include, but are not limited to, those derived from mammalian short and long WAP, alpha, beta, and kappa, casein, alpha and beta lactoglobulin, beta-CN 5' genes, as well as the mouse mammary tumor virus (MMTV) promoter. Such promoters and other elements may be derived from any mammal, including, but not limited to, cows, goats, sheep, pigs, mice, rabbits, and guinea pigs. Promoter and other regulatory sequences, vectors, and other relevant teachings are provided, e.g., by Clark (1998) *J Mammary Gland Biol Neoplasia* 3:337-50; Jost et al. (1999) *Nat. Biotechnol* 17:160-4; U.S. Patent Nos. 5,994,616; 6,140,552; 6,013,857; Sohn et al. (1999) *DNA Cell Biol.* 18:845-52; Kim et al. (1999) *J. Biochem. (Japan)* 126:320-5; Soulier et al. (1999) *Euro. J. Biochem.* 260:533-9; Zhang et al. (1997) *Chin. J. Biotech.* 13:271-6; Rijnkels et al. (1998) *Transgen. Res.* 7:5-14; Korhonen et al. (1997) *Euro. J. Biochem.* 245:482-9; Uusi-Oukari et al. (1997) *Transgen. Res.* 6:75-84; Hitchin et al. (1996) *Prot. Expr. Purif.* 7:247-52; Platenburg et al. (1994) *Transgen. Res.* 3:99-108; Heng-Cherl et al. (1993) *Animal Biotech.* 4:89-107; and Christa et al. (2000) *Euro. J. Biochem.* 267:1665-71; the entire disclosures of each of which is herein incorporated by reference.

In another embodiment, the polypeptides of the invention can be produced in milk by introducing polynucleotides encoding the polypeptides into somatic cells of the mammary gland in vivo, e.g. mammary secreting epithelial cells. For example, plasmid DNA can be infused through the nipple canal, e.g. in association with DEAE-dextran (see, e.g., Hens et al. (2000) *Biochim. Biophys. Acta* 1523:161-171), in association with a ligand that can lead to receptor-mediated endocytosis of the construct (see, e.g., Sobolev et al. (1998) 273:7928-33), or in a viral vector such as a retroviral vector, e.g. the Gibbon ape leukemia virus (see, e.g., Archer et al. (1994) *PNAS* 91:6840-6844). In any of these embodiments, the polynucleotide may be operably linked to a mammary gland specific promoter, as described above, or, alternatively, any strongly expressing promoter such as CMV or MoMLV LTR.

The suitability of any vector, promoter, regulatory element, etc. for use in the present invention can be assessed beforehand by transfecting cells such as mammary epithelial cells, e.g.

MacT cells (bovine mammary epithelial cells) or GME cells (goat mammary epithelial cells), in vitro and assessing the efficiency of transfection and expression of the transgene in the cells.

For in vivo administration, the polynucleotides can be administered in any suitable formulation, at any of a range of concentrations (e.g. 1-500 µg/ml, preferably 50-100 µg/ml), at any
5 volume (e.g. 1-100 ml, preferably 1 to 20 ml), and can be administered any number of times (e.g. 1, 2, 3, 5, or 10 times), at any frequency (e.g. every 1, 2, 3, 5, 10, or any number of days). Suitable concentrations, frequencies, modes of administration, etc. will depend upon the particular polynucleotide, vector, animal, etc., and can readily be determined by one of skill in the art.

In a preferred embodiment, a retroviral vector such as as Gibbon ape leukemia viral vector
10 is used, as described in Archer et al. ((1994) PNAS 91:6840-6844). As retroviral infection typically requires cell division, cell division in the mammary glands can be stimulated in conjunction with the administration of the vector, e.g. using a factor such as estradiol benzoate, progesterone, reserpine, or dexamethasone. Further, retroviral and other methods of infection can be facilitated using accessory compounds such as polybrene.

15 In any of the herein-described methods for obtaining GENSET polypeptides from milk, the quantity of milk obtained, and thus the quantity of GENSET polypeptides produced, can be enhanced using any standard method of lactation induction, e.g. using hexestrol, estrogen, and/or progesterone.

The polynucleotides used in such embodiments can either encode a full-length GENSET
20 protein or a GENSET fragment. Typically, the encoded polypeptide will include a signal sequence to ensure the secretion of the protein into the milk.

Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention:

A further object of the invention comprises recombinant host cells obtained from a
25 transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or a GENSET gene disrupted by homologous recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures
30 with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989), and Shay *et al.* (1991), which disclosures are hereby incorporated by reference in their entireties.

USES OF POLYPEPTIDES OF THE INVENTION

The polypeptides and polynucleotides of the present invention can be used in any of a large
35 number of ways, including numerous in vitro and in vivo uses. Specific uses for many of the herein-described polypeptides and polynucleotides are described in detail below.

Protein of SEQ ID NO:255 (internal designation 500762786_255-24-5-0-A2-R_104)

The cDNA of clone 500762786_255-24-5-0-A2-R_104 (SEQ ID NO:86) encodes the human EDR4 protein

5 LFPAPAPPPAPAFAPPPKVSPERSAPRVPLPSPQPSYPFRPAASGGTPPPACLPPAQPQCQGS
AMNLFRLGDLSHLLAIHLLLLKIWKSRSCAAHPQLPLSFCLSVCLSVLSLSXSLSLSFVSK
KKKK (SEQ ID NO:255). It will be appreciated that all characteristics and uses of the
polynucleotides of SEQ ID NO:86 and polypeptides of SEQ ID NO:355, described throughout the
present application also pertain to the human cDNA of clone 500762786_255-24-5-0-A2-R_104
10 and polypeptides encoded thereby. Polypeptide fragments having a biological activity described
herein and polynucleotides encoding the same are included in the present invention. Related
polynucleotide and polypeptide sequences included in the present invention are SEQ ID NOs:406
and 520.

The normal functioning of the eukaryotic cell requires that all newly synthesized proteins
15 be correctly folded, modified, and delivered to specific inter and extracellular sites. Newly
synthesized membrane and secretory proteins enter a cellular sorting and distribution network
during or immediately after synthesis (cotranslationally or posttranslationally) and are routed to
specific locations inside and outside of the cell. The initial compartment in this process is the
endoplasmic reticulum (ER) where proteins undergo modifications such as glycosylation, disulfide
20 bond formation, and assembly into oligomers. The proteins are then transported through an
additional series of membrane-bound compartments which include the various cisternae of the
Golgi complex, where further carbohydrate modifications occur. Transport between compartments
occurs by means of vesicles that bud and fuse in a specific manner; once within the secretory
pathway, proteins do not have to cross a membrane to reach the cell surface.

25 The complexity of this system has advantages for the cell because it allows proteins to fold
and mature in closed compartments that contain the appropriate enzyme catalysts. It is, however,
dependent on sorting mechanisms that position the enzymes correctly and maintain them in place.

The first organelle in this system, the ER, contains multiple enzymes involved in protein
structure modifications. Among these are BiP (binding protein) which directs the correct folding of
30 proteins and, PDI (protein disulfide isomerase) and a homologue of the 90 kDa heat-shock protein,
both of which catalyze the formation and rearrangement of disulfide bonds (Gething, M. J. and
Sambrook, J. (1992) Nature 355:33-45). These abundant soluble proteins must be retained in the ER
and must be distinguished from the newly synthesized secretory proteins which are rapidly
transported to the Golgi apparatus. The signal for retention in the ER in mammalian cells consists of
35 the tetrapeptide sequence, KDEL, located at the carboxy terminus of proteins. This sequence was
first identified when the sequences of rat BiP and PDI were compared and it was subsequently
found at the carboxy terminus of other luminal ER proteins from a number of species (Munro, S.
(1986) Cell 46:291-300; Pelham, H. R. (1989) Ann. Rev. Cell. Biol. 5:1-23). Proteins containing

this sequence leave the ER but are quickly retrieved from the early Golgi compartment and returned to the ER, while proteins without this signal continue through the distribution pathway.

Two endoplasmic retrieval receptors were first identified in *S. cerevisiae*; two human endoplasmic retrieval receptors were subsequently isolated by the use of degenerate PCR primers based on the *S. cerevisiae* sequences (Hardwick, K. G. (1990) EMBO J. 9:623-630; Semenza, J. C. (1990) Cell 61:1349-1357; Lewis, M. J. and Pelham, H. R. (1990) Nature 348:162-163; Lewis, M. J. and Pelham, H. R. (1992) J. Mol. Biol. 226:913-916). Comparisons of these sequences shows that they consist of a conserved 7-transmembrane domain structure with only short loops in the cell cytoplasm and the ER lumen. Studies with these endoplasmic retrieval receptors show that ligand binding controls the movement of the receptor; when expressed in COS cells, the human receptor is normally concentrated in the Golgi, but moves to the ER when bound to a ligand such as KDEL-tagged hen lysozyme (Lewis, M. J. and Pelham, H. R. (1992) Cell 68:353-364).

The ER retrieval function of these molecules serves to maintain the pool of enzymes in the ER that are necessary to perform protein structure modifications, retains newly synthesized proteins in the ER until they have been correctly modified, and regulates the structure of the Golgi apparatus. *Saccharomyces cerevisiae* cells that lack an ER retrieval receptor (Erd2) have a defective Golgi apparatus and fail to grow. Analysis of yeast Erd2 mutants suggests that their growth requires both the retention of multiple proteins in the ER and the selective removal of specific proteins from the Golgi (Townsend, F. M. (1994) J. Cell Biol. 127:21-28). Overexpression of a human ER retrieval receptor in COS cells results in hyperactive retrograde traffic from the Golgi to the ER leading to a loss of the Golgi structure and the breakdown of the secretory pathway (Hsu V. W. (1992) Cell 69:625-635).

Disruptions in the cellular secretory pathway have been implicated in several human diseases. In familial hypercholesterolemia the low density lipoprotein receptors remain in the ER, rather than moving to the cell surface (Pathak, R. K. (1988) J. Cell Biol. 106:1831-1841). A form of congenital hypothyroidism is produced by a deficiency of thyroglobulin, the thyroid prohormone. In this disease the thyroglobulin is incorrectly folded and is therefore retained in the ER (Kim, P. S. (1996) J. Cell Biol. 133:517-527). Mutant forms of proteolipid protein (PLP) have been examined as they play a role in generating dysmyelinating or hypomyelinating diseases. In this case, the mutations that result in disease are mutations that arrest transport of PLP in the ER and the early Golgi; the subsequent accumulation of PLP in the ER results in rapid oligodendrocyte death (Gow, A. (1994) J. Neurosci. Res. 37:574-583).

The human ER retrieval receptor function is necessary for processing and presentation of specific antigens to T cells. Many antigens must be processed intracellularly before they can be presented, in association with major histocompatibility complex (MHC) molecules at the cell surface, for recognition by the antigen-specific receptor of T cells. Disruption of the ER retrieval receptor function with an antibiotic, Brefeldin A, abolishes the ability of a cell to present these specific antigen complexes to T cells. These antigenic proteins must be retained in the ER for

cleavage to smaller peptides which can then bind to MHC molecules and be released for presentation at the cell surface. (Kakiuchi, T. (1991) J. Immunol. 147:3289-3295).

The discovery of polynucleotides encoding a novel human KDEL receptor, and the molecules themselves, provides the means to further investigate the regulation of the cellular protein secretory pathway. Discovery of molecules related to a novel human KDEL receptor satisfies a need in the art by providing a means or a tool for the study of this pathway and the diseases that involve the dysfunction of this pathway.

In an embodiment of the present invention, ERD4 polypeptides of the present invention are used to purify KDEL containing proteins and other homologous proteins with similar signals for ER retention such as "HDEL", "DDEL", "ADEL", "SDEL", "RDEL", "KEEL", "QEDL", "HIEL", "HTEL" and "KQDL". This may be carried out by covalently or non-covalently attached the ERD4 polypeptides of the present invention to a column or other solid support using techniques well known in the art (e.g., affinity chromatography, panning, etc.). Once bound to the ERD4 polypeptide, the complex is washed to remove contaminants. The target protein is released using increasing salt concentrations either in a gradient or step type purification. The bound target protein may also be released from the ERD4 polypeptide by a single step up in salt concentration.

In another embodiment of the present invention, the EDR4 polypeptides of the present invention are used to detect KDEL containing proteins and other homologous sequences as described above by methods comprising the steps of contacting KDEL or other homologous sequences with an EDR4 polypeptide under conditions that allow binding to said sequence, and detecting the presence of bound EDR4. The presence of bound EDR4 can be detected using methods known in the art, such as by labeling EDR4 directly or indirectly. Bound EDR4 can be detected, for example, by using an antibody that specifically binds to EDR4 or another EDR4 - binding compound that is detectable directly or indirectly.

Preferred ERD4 polypeptides for binding KDEL containing proteins and other homologous sequences described above comprise the amino acid sequence -KIWK- or -MNLFRFLGDLSHLLAIHLLLLKIWKSRSCA-.

The present invention is further directed to a transformant comprising the following expression units in a co-expressible state: an expression unit containing a gene coding for an ERD4 polypeptide which is capable of binding to a protein localizing in the endoplasmic reticulum and having a signal for staying therein; an expression unit containing a gene coding for said protein localizing in endoplasmic reticulum; and an expression unit containing a foreign gene coding for a polypeptide which is a subject of function of said protein localizing in endoplasmic reticulum, and to a transformant comprising, in a co-expressible state, a fusion gene which is composed of a DNA fragment coding for a human serum albumin prepro-sequence and a foreign gene coding for a useful polypeptide. The present invention is also directed to a process for producing said polypeptide by co-expressing said genes in said transformant such that the polypeptide is predominantly secreted out of the transformant cell. Consequently, the invention has an advantage of improving the productivity of said polypeptide.

More particularly, the invention relates to: A transformed yeast cell comprising the following expression units integrated on a yeast chromosome in a co-expressible state: a first expression unit containing a gene coding for a receptor for an endoplasmic reticulum retention signal, wherein the receptor is the receptor protein ERD4 or a fragment thereof which is capable of binding to a retention signal selected from the group consisting of "KDEL", "HDEL", "DDEL", "ADEL", "SDEL", "RDEL", "KEEL", "QEDL", "HIEL", "HTEL" and "KQDL". and a second expression unit containing a gene encoding a protein disulfide isomerase, wherein said isomerase comprises an endoplasmic reticulum retention signal, or a gene encoding a fusion protein comprising the amino acid sequence of said isomerase and a human serum albumin prepro-sequence. These methods can be carried out using methods known in the art or described in U.S. Patent 5,578,466, incorporated herein by reference in its entirety.

Proteins of SEQ ID NO:193 and 194 (internal designation 585770_215-16-5-0-E8-F and 123996_140-002-5-0-B4-F)

The cDNA of clones 585770_215-16-5-0-E8-F (SEQ ID NO:24) and 123996_140-002-5-0-B4-F (SEQ ID NO:25) encode the human Smooth Muscle and Pain Effector (SMPE) proteins: MRGATRVSIMLLLVTVSDCAVITGACERDVQCGAGTCCAISLWLRGLRMCTPLGRXGEEC HPGSHKIPFFRKRKHHTCPCLPNLLCSRFPDGRYRCSMDLKNINF (SEQ ID NO:193) and MRGATRVSIMLLLVTVSDCAVITGACERDVQCGAGTCCAISLWLRGLRMCTPLGREGEEC HPGSHKIPFFRKRKHHTCPCLPNLLCSRFPDGRYRCSMDLKNINF (SEQ ID NO:194), respectively. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:24 and 25 and polypeptides of SEQ ID NO:193 and 194, described throughout the present application also pertain to the human cDNA of clones 585770_215-16-5-0-E8-F and 123996_140-002-5-0-B4-F, and the polypeptides encoded thereby. Polypeptide fragments having a biological activity described herein and polynucleotides encoding the same are also included in the present invention. Related polynucleotide and polypeptide sequences included in the present invention are SEQ ID NOs:360 and 447.

SMPE contracts longitudinal ileal muscle and distal colon, and relaxes the proximal colon. SMPE binds with a high affinity to both ileum and brain membranes. Therefore, included as embodiments of the present invention is a method of causing gastrointestinal smooth muscle cells to contract, in vitro or in vivo, comprising the steps of contacting said cells with a contracting effective amount of an SMPE polypeptide. Preferably, the gastrointestinal smooth muscle cells are those of the longitudinal ileal or distal colon. A further embodiment of the present invention is a method of causing gastrointestinal smooth muscle cells to relax comprising the steps of contacting said cells with a relaxing effective amount of an SMPE polypeptide. Preferably, the gastrointestinal smooth muscle cells are proximal colon cells. SMPE can also be used in the same manner to contract uterine cells. Therefore, included in the present invention is a method of causing uterine smooth muscle cells to contract comprising contacting said cells with a contracting effective amount of an

SMPE polypeptide. Further included in the present invention is a method of causing smooth muscle cells (e.g., bladder, vascular) to contract comprising contacting said cells with a contracting effective amount of an SMPE polypeptide. Further included in the present invention is a method of inhibiting angiogenesis comprising contacting vascular endothelial cells with an angiogenesis inhibiting effective amount of an SMPE polypeptide. The SMPE anti-angiogenic effect can be measured using assays known in the art. For example, the anti-angiogenic effect in vivo can be assayed by using the 10-day-old embryo chick chorioallantoic membrane model.

SMPE binds with a high affinity to both ileum and brain membranes. Therefore, as a further embodiment of the present invention is a method of binding an SMPE polypeptide to ileum or brain membranes. The method can be further used as a method of detecting ileum or brain membranes comprising the steps of contacting ileum or brain membranes with an SMPE polypeptide under conditions that allow binding to said membranes, and detecting the presence of SMPE. The presence of SMPE can be detected using methods known in the art, such as by labeling SMPE directly or indirectly. Bound SMPE can be detected, for example, by using an antibody that specifically binds to SMPE or another SMPE-binding compound that is detectable directly or indirectly.

SMPE is also expressed in spermatocytes. Therefore, a further embodiment of the present invention is a method of detecting testes or spermatocytes by detecting an SMPE polypeptide or nucleic acid. An SMPE polypeptide can be detected using anti-SMPE antibodies or other SMPE-binding compounds. SMPE polynucleotides, such as mRNA, can be detected using methods known in the art such as PCR (RT-PCR), hybridization (Northern blot analysis), etc.

SMPE elicits hyperalgesia when it contacts the CNS, e.g., the brain. Therefore, the present invention includes a method of causing hyperalgesia comprising contacting the CNS with a hyperalgesia effecting amount of an SMPE polypeptide. SMPE can be delivered to the CNS using methods well known in the art including those described in PCT application WO9906060, incorporated herein by reference in its entirety. Using the methods of WO9906060, the TGF- α or other polypeptide that binds the epidermal growth factor (EGF) receptor, is substituted with an SMPE polypeptide of the present invention.

Further included in the present invention are methods of inhibiting the above SMPE activities using an inhibitor of SMPE. A preferred inhibitor of SMPE is an anti-SMPE antibody. Thus, an embodiment of the present invention is a method of inhibiting smooth muscle contraction (bladder, gastrointestinal cells, uterine) or pain comprising the step of contacting said cells with an effective contractive or pain inhibiting amount of an anti-SMPE antibody or other SMPE inhibitor.

The invention further relates to a method of screening for test compounds that bind and/or inhibit an SMPE activity above comprising the steps of contacting an SMPE polypeptide with said test compound and detecting or measuring whether said test compound binds said SMPE polypeptide. Alternatively, the method comprises the steps of contacting an SMPE polypeptide with a binding target (e.g., smooth muscle cells or brain cells) of said SMPE polypeptide in the presence of a test compound, and detecting or measuring the binding of the SMPE polypeptide to

said binding target, wherein a difference in the amount of said binding in the presence of said test compound relative to the amount of binding in the absence of the test compound indicates that the test compound modulates, preferably inhibits, the binding of said polypeptide to said binding target. The method may alternatively comprise the steps of contacting an SMPE polypeptide with a

5 binding target in the presence of a test compound, wherein the binding of said SMPE polypeptide with said binding target elicits or causes a biological activity (e.g., activities described above) which is detected or measured, and further wherein a difference in the level of said biological activity in the presence of the test compound relative to the amount of biological activity in the absence of the test compound indicates that the test compound modulates, preferably inhibits or activates, the

10 biological activity of said SMPE polypeptide.

Preferred SMPE polypeptides for use in the methods described herein include the amino acid sequences –

AVITGACERDVQCGAGTCCAISLWLRGLRMCTPLGREGEECHPGSHKIPFRKRKHH- or -

TGACERDVQCGAGTCCAISLWLRGLRMCT- of SEQ ID NO:193 or 194.

15

Protein of SEQ ID NO:305 (internal designation 500691428_255-2-5-0-D4-R_104)

The human cDNA of clone 500691428_255-2-5-0-D4-R_104 (SEQ ID NO:136) encodes the human VESICLE-ASSOCIATED MEMBRANE PROTEIN 10 or VAMP-10 protein:

MSATAATAPPAAPAGEGGPPAPPPNLTSNRRLLQQTQAQVDEVVDIMRVNVDKVLERDQKL

20 SELDDRADALQAGPSQFETSAAKLKRKYWWKNLKMMLGVICAILIIIVYFST (SEQ ID NO:305). It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:136 and polypeptides of SEQ ID NO:305 described throughout the present application also pertain to the human cDNA of clone 500691428_255-2-5-0-D4-R_104 and the polypeptides encoded thereby. Polypeptide fragments having a biological activity described herein and

25 polynucleotides encoding the same are also included in the present invention. Related polynucleotide and polypeptide sequences included in the present invention are SEQ ID NOs:432 and 546.

VAMP-10 is an integral membrane protein involved in the movement of vesicles from the plasmalemma of one cell, across the synapse, to the plasma membrane of the receptive neuron. This

30 regulated vesicle trafficking pathway and the endocytotic process may be blocked by the highly specific action of clostridial, tetanus toxin (TeTx) and botulinum toxin (BoNT) and other metalloendoprotease neurotoxins which prevents neurotransmitter release by cleaving VAMPs. VAMP-10 is important in membrane trafficking. It participates in axon extension via exocytosis during development, in the release of neurotransmitters and modulatory peptides, and in

35 endocytosis. The tightly-regulated synaptic vesicle cycle at the nerve terminal consists of the formation of synaptic vesicles, the docking of vesicles comprising VAMP-10 to the presynaptic plasma membrane, the fusion of these membranes and consequent neurotransmitter release, endocytosis of the empty vesicles and the regeneration of fresh vesicles. Endocytotic vesicular

transport includes such intracellular events as the fusions and fissions of the nuclear membrane, endoplasmic reticulum, Golgi apparatus, and various inclusion bodies such as peroxisomes or lysosomes.

VAMP-10, like other VAMPs, has a three domain organization. The domains include a
5 variable proline-rich, N-terminal sequence, a highly conserved central hydrophilic core of amino acids, and a hydrophobic sequence of amino acids presumed to be the membrane anchor.

In one aspect, the invention includes a VAMP-10 polypeptide composition for use in delivering a second composition, preferably nucleic acids, polypeptides, or small molecules such as therapeutic drugs, to target biological cells either in vitro or in vivo. The composition comprises a
10 VAMP-10 polypeptide as a first molecule and a second molecule. The second molecule may, if desirable, be covalently or non-covalently attached or fused to the VAMP-10 polypeptide. The VAMP-10 polypeptide composition may further comprise artificial lipids to facilitate delivery of the second molecule by liposomes or lipid vesicles. Methods for using VAMP-10 polypeptides in these methods are known in the art and include U.S. Patents 6,074,844, 6,203,794 and 6,099,857,
15 incorporated by reference in their entireties. In a preferred embodiment, VAMP-10 polypeptides are used to facilitate delivery of a second composition, e.g., liposome mediated DNA transfection, to cells in culture, preferably neuronal cells, and further preferably to the presynaptic membrane.

VAMP-10 polypeptides are also useful in methods of inhibiting the release of neurotransmitters by preventing the docking and/or fusing of a presynaptic vesicle to the
20 presynaptic membrane. These polypeptides may be referred to as excitation-secretion uncoupling peptides (ESUPs). Fragments of VAMP-10 having this blocking activity can be identified using methods known in the art (See e.g., U.S. Patents 6,090,631 and 6,169,074 incorporated by reference in their entireties). ESUPs of the present invention comprise synthetic and purified VAMP-10 peptide fragments which correspond in primary structure to peptides which serve as binding
25 domains for the assembly of a ternary protein complex ("docking complex") which is critical to neuronal vesicle docking with the cellular plasma membrane prior to neurotransmitter secretion. Preferably, the primary sequence of the ESUPs of the invention also includes amino acids which are identical in sequence to the VAMP-10 peptide products of BoTx and TeTx proteolytic cleavage in neuronal cells, or fragments thereof ("proteolytic products"). For optimal activity, ESUPs of the
30 invention have a minimum length of about 20 amino acids and a maximal length of about 28 amino acids, although they may be larger or smaller. Preferably, the ESUPs correspond in primary structure to binding domains in the docking complex, most preferably the region of such binding domains that are involved in the formation of a coiled-coil structure in the native docking complex proteins. ESUPs may also be used as pharmaceutical carriers as part of fusion proteins to deliver
35 substances of interest into neural cells in a targeted manner. Preferred VAMP-10, or ESUP, polypeptides for use in inhibiting the release of neurotransmitters include those comprising –
NRRLQQTQAQVDEVVDIMRVNVDKVLERDQKLSELDDRADALQAGPSQFETSAAKLKRK
- of SEQ ID NO:305. More preferred ESUP polypeptides comprise an amino acid sequence portion of SEQ ID NO:305 selected from the group consisting of: RVNVDKVLERDQKLSELDD;

KVLERDQKLSELDDRA; VNVDKVLERDQKLSELDDRA;
 DIMRVNVDKVLERDQKLSELDDRADAL; DEVVDIMRVNVD; QAQVDEVVDIMRVNVD;
 LQQTQAQVDEVVDIMRVNVD; QQTQAQVDEVVD; NRRLQQTQAQVDEVVD; and
 NLTSNRRLQQTQAQVDEVVD.

5 The ESUPs above may be used to inhibit or treat pain according to U.S. Patent 6,113,915 or 5,989,545 (incorporated by reference herein in their entireties) by substituting the polypeptides of the present invention for BoTx type A.

10 Because VAMP-10 is a component of vesicles, antibodies to VAMP-10 are useful in the detection of vesicles, preferably neuronal vesicles transporting neurotransmitters. VAMP-10 can be used during purification of vesicles as a marker for vesicles or vesicles can be detected using antibodies to VAMP-10 in assays such as immunohistochemistry. Following exocytosis of vesicles, a portion of the VAMP-10 inserted in the vesicle appears on the surface of the axon, thus making VAMP-10 useful for the detection and monitoring of exocytosis of synaptic vesicles.

15 Detection of VAMP-10 expression (mRNA or protein) levels or mutated forms of VAMP-10 is further useful in the determination or diagnosing of whether someone is at risk of developing or has a neurological disorder, such as mood disorders selected from depression, bipolar disorder, schizophrenia, etc.), wherein a decreased level in expression of VAMP-10, mRNA or protein, as compared to an individual without a neurological disorder indicates the individual has the disorder or is at risk of having the disorder in the future.

20 The present invention further includes a novel assay system for toxins, such as clostridial, tetanus toxin (TeTx) and botulinum toxin (BoNT), using novel reagents. Preferably, methods of U.S. Patent 6,043,042, incorporated by reference in its entirety, are used to perform the assay, wherein a VAMP-10 polypeptide is the substrate cleaved by the test compound. More specifically, the assay comprises the steps of:

25 The invention relates to an assay for botulinum toxin or tetanus toxin comprising the steps of:

30 (a) combining a test compound with a substrate and with antibody, wherein the substrate has a cleavage site for the toxin and when cleaved by toxin forms a product, and wherein the antibody binds to the product but not to the substrate; and wherein the substrate is a VAMP-10 polypeptide; and

 (b) testing for the presence of antibody bound to the product, which product is attached to a solid phase assay component.

35 Preferably, in the practice of this invention, the VAMP-10 polypeptide is cleaved by the toxin to generate new peptides having N- and C-terminal ends. In addition, the peptide substrate is attached to a solid phase component of the assay.

 The assay according to the invention may utilize assay components (a) and (b):

 (a) a peptide linked to a solid-phase, the peptide being cleavable by the toxin to generate a cleavage product,

(b) an antibody that binds to the cleavage product but not to the uncleaved polypeptide or an antibody that binds a cleavage product that is either the N-terminal or C-terminal portion of the VAMP-10, and the assay may comprise the steps of:

- (i) combining a test compound that may contain or consist of the toxin with the solid-phase peptide to form an assay mixture,
- (ii) subsequently or simultaneously combining the assay mixture with the antibody, and
- (iii) subsequently or simultaneously determining whether there has been formed any conjugate between the antibody and the cleavage product.

Preferably, the step (i) of the assay is carried out in the presence of a zinc compound and a VAMP-10 polypeptide.

In this embodiment, the assay comprises:

- (i) combining the test compound with a solid phase comprising a VAMP-10 polypeptide,
- (ii) washing the test compound from the solid phase,
- (iii) combining the solid phase with an antibody adapted for binding selectively with peptide cleaved by toxin, and
- (iv) detecting a conjugate of the antibody with cleaved peptide.

In another embodiment, the assay comprises:

- (i) adding a test solution to an assay plate comprising immobilized peptide, the peptide being a VAMP-10 polypeptide;
- (ii) incubating the assay plate,
- (iii) washing the plate with a buffer,
- (iv) adding to the plate an antibody solution, said solution comprising an antibody adapted selectively to bind to a peptide selected from the group consisting of (I) the 50 C-terminal amino acid residues SEQ ID NO:305, the 30 C-terminal amino acid residues SEQ ID NO:305, and the 20 C-terminal amino acid residues SEQ ID NO:305 (any other VAMP-10 polypeptide of the present invention may also be selected).
- (2) a peptide the N-terminal end of which is selected from the group consisting of: (I) the 50 N-terminal amino acid residues SEQ ID NO:305, the 30 N-terminal amino acid residues SEQ ID NO:305, and the 25 N-terminal amino acid residues SEQ ID NO:305 (any other VAMP-10 polypeptide of the present invention may also be selected).
- (v) incubating the assay plate,
- (vi) washing the plate with a buffer, and
- (vii) measuring the presence of antibody on the assay plate.

In this embodiment, the antibody may be linked to an enzyme and the presence of antibody on the plate is measured by adding an enzyme substrate and measuring the conversion of the substrate into detectable product. The detectable product may be colored and measured by absorbance at a selected wavelength.

In the practice of the invention, the inactive toxin present in the test compound may be converted to active toxin. This may be accomplished by adding a protease to the test compound.

The antibody-peptide conjugate may be detected using a further antibody specific to the first antibody and linked to an enzyme.

5 Proteins of SEQ ID NO:171 (internal designation Clone ID:589115) and related protein of SEQ ID NO:457.

The polynucleotides of SEQ ID NO:2 and SEQ ID NO:340 and polypeptides of SEQ ID NO:171 and 457 encode a C-terminal variant of Apolipoprotein A1, herein referred to as ApoAI-CTV. An embodiment of the invention includes compositions of SEQ ID NO:2, 340, 171, and 457 which encode for this novel variant of the apolipoprotein family of lipid transporting proteins. Specifically, ApoAI-CTV is a component of high density lipoprotein which functions to remove cholesterol from circulation and thus providing protection against the development of atherosclerosis, coronary atherosclerotic lesions and subsequent microvascular and cardiovascular disease.

15 Preferred polynucleotides of the invention are compositions of the novel portion of the cDNA from bases 465 to 521 of SEQ ID NO:2 including the nucleic acids comprising the sequences –

GCAGCTTTCTTAACTATCCTAACAAGCCTTGGACCAAATGGAAATAAAGCTTTTTGA-, -
GAAGGCAGCTTTCTTAACTATCCTAACAAGCCTTGGACCAAATGGAAATAAAGCTTTTT
20 GA-, or –

AGCTCTACCGCCAGAAGGCAGCTTTCTTAACTATCCTAACAAGCCTTGGACCAAATGG
AAATAAAGCTTTTTGATGAAAAA- of SEQ ID NO:2 and 340.

Preferred polypeptides of the invention are compositions of the novel C-terminal portion comprising the amino acid sequence -AAFLTILTS LGPNGNKAF, -

25 MELYRQKAAFLTILTS LGPNGNKAF, or -QKKWQEEMELYRQKAAFLTILTS LGPNGNKAF of SEQ ID NO:171 and SEQ ID NO:457.

Further preferred polypeptides of the invention include the compositions comprising the apolipoprotein domain

KA AVLTLAVLFLTGSQARHFWQDEPPQSPWDRVKDLATVYVDVLKDSGRDYVS

30 QFEGSALGKQLNLKLLDNWDSVTSAFSNLREQLGPVTQEXWDNLEKETEGLRQEMSKDLE
EVKAKVQPYLDDFQKKWQEEMELYRQKAAFLTILTS LGPNGNKA of SEQ ID NO:171 and 457, or the amino acid residue positions –17 to +141 of SEQ ID NO:171.

An embodiment of the invention includes a method for treatment of atherosclerosis or cardiovascular diseases, comprising administering to an individual a therapeutically effective

35 amount of apoAI-CTV or variants or mixtures thereof to lower total plasma cholesterol at least 5% of pretreatment levels.

Further utility of the polypeptides of the present invention may be further confirmed by methods of production and use of other apolipoproteins by those skilled in the art or as described by

Ageland et al in US Patent 5990081, which disclosure is hereby incorporated by reference in its entirety.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:2 and SEQ ID NO:340 and polypeptides of SEQ ID NO:171 and 457 described throughout the present application also pertain to the human cDNA of clone 589115.

Proteins of SEQ ID NO:302 (internal designation Clone ID:1000853793) and related protein of SEQ ID NO:543.

MRLFLSLPVLVVVLSIVLEGPAPAQGTPDVSSALDKLKEFGNTLEDKARELISRIKQ
10 SELSAKMREWFSETFQKVKDKLKIDS

The polynucleotides of SEQ ID NO:133 and SEQ ID NO:429 encode human apolipoprotein CI (ApoCI) polypeptide of SEQ ID NO:302 and SEQ ID NO:543, respectively. The ApoCI of the invention differs by 1 amino acid comprising the amino acid sequence FQKVKDKLKI, where aspartate (D at position 77 of SEQ ID NO:302) replaces a glutamate (E) of the ApoCI of GENPEP
15 accession X00570, AF050154, and M20902. ApoCI is a member of the apolipoprotein family of lipid binding and transporting proteins specifically functioning to transport cholesterol esters.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:133 and SEQ ID NO:429 and polypeptides of SEQ ID NO:302 and SEQ ID NO:543, described throughout the present application also pertain to the human cDNA of clone 1000853793, and the
20 polypeptides encoded thereby.

Proteins of SEQ ID NO:295 (internal designation Clone ID:642948), SEQ ID NO:296 (internal designation Clone ID:638743), and SEQ ID NO:539.

MEASALTSSAVTSVAKVVRVASGSAAVVLPLARIATVVIGGVVAMAAVPMVLSAMGFTAA
25 GIASSSIAAKMMSAAAIANGGGVASGSLVATLQSLGATGLSGLTKFILGSIGSAIAAVIARFY

The polynucleotides of SEQ ID NO:126, 127, and 425 and the polypeptides of SEQ ID NO:295, 296 and 539 encode human transmembrane, alpha-interferon-inducible polypeptides, aINFIP-1, aINFIP-1, and aINFIP-3, respectively. Preferred polynucleotides and polypeptides of the invention comprise the nucleic acid sequences of SEQ ID NO:126, 127, and 425 and amino acid
30 sequences of SEQ ID NO:295, 296 and 539.

Preferred polypeptides of SEQ ID NO:295 and SEQ ID NO:539 for use in the methods described herein include the amino acid sequences comprising –
VLSAMGFTAAGIASSSIAAKMMSAAAIANGGGVASG-, -
SSIAAKMMSAAAIANGGGVASGSLVATLQSLGAT-, or -
35 VIGGVVAMAAVPMVLSAMGFTAAGIASSSIAAKMMSAAAIANGGGVASGSLVATLQSLG
ATGLSGLTK-.

Preferred polypeptides of SEQ ID NO:296 for use in the methods described herein include the amino acid sequence –AAAIANXGGVASGSLVATLQSLGATGLSGLTKF- or -LSAMGFTAAGIASSSIAAKMMSAAAIANXGGVASGSLVATLQSLGATGLS-.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:126, 127, and 425 and the polypeptides of SEQ ID NOs:295, 296 and 539, described throughout the present application also pertain to the human cDNA of Clone ID:642948 and Clone ID:638743, and the polypeptides encoded thereby.

Sites of glycine myristylation within a polypeptide function to modulate the activity and compartmentalization of the protein (Resh, M. D. Biochim Biophys Acta 1451:1-16 (1999)).

Preferred polypeptides of the invention include fragments comprising the sites of N-myristylation. Preferred amino acids of said sites within SEQ ID NO:295 and SEQ ID NO:539 include GGVMAM (positions 39-44), GIASSS (positions 60-65), GGGVAS (positions 79-84), GSLVAT (positions 85-90), and GSIGSX (positions 108-113). Further preferred are amino acids within 6 residues preceding or 6 residues following said amino acid sequences. Further preferred amino acids include sequences comprising the sites of N-myristylation in the polypeptides of SEQ ID NO:296. Preferred amino acids of said sites within SEQ ID NO:296 include IATVVIGGVVAMAAVPMV, MGFTAAGIASSSIAAKMM, AAIANXGGVASGSLVATL, NXGGVASGSLVATLQSLGA, and LTKFILGSIGSAIAAVIAR.

Interferons (IFNs) are a part of the group of intercellular messenger proteins known as cytokines and are part of the body's natural defense to viruses and tumors. Type I IFNs (alpha and beta interferons) are produced in a variety of cells types and their biosynthesis is stimulated by viruses and other pathogens, and by various cytokines and growth factors. Both α - and γ -IFNs are immunomodulators and anti-inflammatory agents, activating macrophages, T-cells and natural killer cells (reviewed in Jonasch and Haluska, Oncologist 6(1):34 (2001)). As part of the body's natural defense to viruses and tumors, IFNs affect the function of the immune system and have direct action on pathogens and tumor cells. IFNs mediate these multiple effects by inducing the synthesis of cellular proteins, including the polypeptides of the present invention, aINFIP-1, aINFIP-1, and aINFIP-3.

Antiviral activity of the aINFIP polypeptides are assayed according to conventional methods (Tovey et al, Proc. Soc. Exp. Biol. and Med., 1974 146: 809-815). Preferred polypeptides of SEQ ID NO:295, 296 and 539 and fragments thereof include those which possess antiviral function, where preferred antiviral activity is against herpes simplex virus and hepatitis virus C, alone or in combination with known antiviral treatments such as interferon alpha.

The antitumor activity the aINFIP polypeptides of the invention can be demonstrated by similar methods using tumor cell lines rather than treatment of cells with virus as used to test antiviral activity. Tumor cell lines examples include MCF-7 (human breast cancer derived), NOS-1 (human oral primary squamous cell carcinoma derived), and MedB-1 (human primary mediastinal large B-cell lymphoma derived).

Further utility of the polypeptides of the present invention may be further confirmed by methods of interferon inducible proteins in the inhibition of viral functions such as cell penetration, uncoating, RNA and protein synthesis, assembly and release described in Hardman et al., Pharmacological Basis of Therapeutics, McGraw-Hill, New York NY pp 1211-1214, 25 (1996),
5 disclosure of which is hereby incorporated by reference in its entirety.

Another embodiment of the present invention relates to the use of aINFIP polypeptides or fragments thereof to treat and/or prevent the ill-effect of bacterial infection. In a preferred embodiment, the protein of the invention may be used to counteract the effects of the bacterial endotoxin lipopolysaccharide (LPS). The methods for using such compositions is described in
10 Dziegielewska and Andersen, Biol. Neonate, 74:372-5 (1998), the disclosure of which is incorporated herein by reference in its entirety.

Furthermore, the aINFIP polypeptides or fragments thereof may be used to identify specific molecules with which it binds such as agonists, antagonists or inhibitors. Another embodiment of the present invention relates to methods of using the aINFIP polypeptides or fragments thereof to
15 identify and/or quantify cytokines of the interferon family as well as other cytokines such as IL10 and tumor antigens, which may interact with the aINFIP polypeptides of the invention.

The aINFIP polypeptides of the invention or fragments thereof are included in pharmaceutical preparations for treatment, prevention or alleviation of cancers. In another embodiment of the present invention, the aINFIP polypeptides of the invention or fragments thereof
20 are used included in pharmaceutical preparations for treatment, prevention or alleviation of viral or bacterial infections. In another embodiment of the present invention, the aINFIP polypeptides of the invention or fragments thereof are used to inhibit and/or modulate the effect of cytokines and related molecule such as Il-2, TNF alpha, CTLA4, CD28, and others, by preventing the binding of the endogenous cytokine to their natural receptors, thereby blocking cell proliferation or inhibitory
25 signals generated by the ligand-receptor binding event.

In another embodiment of the present invention, the aINFIP polypeptides of the invention or fragments thereof are useful to correct defects in in vivo models of disease such as autoimmune, inflammation and tumor models, by injecting the protein either intra peritoneally intravenously, subcutaneously or directly in the diseased tissue.

30 The polynucleotides of SEQ ID NO:126, 127, and 425 or fragments thereof is useful in diagnostic assays for aINFIP-1, aINFIP-2, or aINFIP-3 gene expression in in vitro models or in conditions associated with expression of the aINFIP polypeptides of the invention. The diagnostic assay is useful to distinguish between absence, presence, and excess expression of the gene and to monitor regulation of levels of the gene of the invention during therapeutic intervention. The DNA
35 may also be incorporated into effective eukaryotic expression vectors and directly targeted to a specific tissue, organ, or cell population for use in gene therapy to treat the above mentioned conditions, including tumors and/or to correct disease- or genetic-induced defects in any of the above mentioned proteins including the protein of the invention.

Protein of SEQ ID NO:170 (internal designation Clone ID:502084) and related protein of SEQ ID NO:456.

The polynucleotides of SEQ ID NO:339 and polypeptides of SEQ ID NO:456 encode neutrophil stimulating protein 2, previously described in WO 9006321 (GENPEP accession A01319) as a novel factor having neutrophil-stimulating activity. The polynucleotide of SEQ ID NO:1 encodes a novel polypeptide variant, neutrophil stimulating protein 2v, comprising the amino acid sequence of SEQ ID NO:170 in which an aspartate (D) residue is located at position +16 of SEQ ID NO:170 rather than a glutamate (E). Preferred compositions of the invention include the polypeptides of SEQ ID NO:170. Further preferred amino acids of SEQ ID NO:170 comprise the sequence LAKGKDESLDS, QXKRNLAKGKDESLDSPLYAE, or SSTKGQXKRNLAKGKDESLDSPLYAELRCMCIKTTSGIHPKNIQSLEVIGKGTHCNQVEVIATLKDGRKICLDPDAPRIKKIVQKKLAGDESAD. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:1 and SEQ ID NO:339 and polypeptides of SEQ ID NO:170 and SEQ ID NO:456, described throughout the present application also pertain to the human cDNA of clone 502084, and the polypeptides encoded thereby.

A preferred embodiment of the invention includes use of the novel neutrophil stimulating protein 2v of SEQ ID NO:170 in a method to stimulate wound healing by contacting the wound area with effective amount of polypeptide of SEQ ID NO:170 or further use as described in US Patent 5,804,176, which disclosure is hereby incorporated by reference in its entirety. A further preferred includes use of neutrophil stimulating protein 2v in the enhancement of angiogenesis for revascularization after injury such following myocardial infarction, wherein site of injury is contacted with effective amount of polypeptide of SEQ ID NO:170 or use as further described in US Patent 5,871,723, which disclosure is hereby incorporated by reference in its entirety. Antibodies against neutrophil stimulating protein 2v, by preventing or blocking the deposition of connective tissue matrix, are useful in the treatment of fibrotic disorders by contacting the polypeptides of SEQ ID NO:170 with fibrotic tissue, such as in scleroderma, liver cirrhosis, and myelofibrosis.

An embodiment of the invention includes fragments of SEQ ID NO:170 which comprise domains which impart function to this cytokine. Preferred fragments include the amino acid sequence comprising the IL8 domain, DSDLYAELRCMCIKTTSGIHPKNIQSLEVIGKGTHCNQVEVIATLKDGRKICLDPDAPRIKKIVQKKL. Further preferred amino acids include the small cytokines (intercrine/chemokine) C-x-C subfamily signature of the amino acid sequence comprising CMCIKTTSGIHPKNIQSLEVIGKGTHCNQVEVIATLKDGRKICLD.

Further preferred polypeptides include portions comprising sites of Protein Kinase C phosphorylation including amino acid residues 2 to 4, residues 13 to 15, residues 36 to 38 and residues 97 to 99 of SEQ ID NO: or amino acids sequence comprising SLR, SAR, STK, and TLK. Further preferred polypeptides include portions of the amino acid sequence comprising sites of

Casein kinase II phosphorylation including amino acid residues 97 to 100 or the amino acid sequence comprising TLKD.

Further preferred polypeptides include portions of the amino acid sequence comprising sites of N-myristylation or the amino acid residues comprising GTHCNQ.

- 5 Further preferred polypeptides include the small cytokines (intercrine/chemokine) C-x-C subfamily signature of the amino acid sequence comprising
CMCIKTTSGIHPKNIQSLEVIGKGTHCNQVEVIATLKDGRKICLD.

Proteins of SEQ ID NO: 227 (internal designation Clone ID: 166601) and related protein of SEQ ID

- 10 NO: 502.

Polynucleotides of SEQ ID NO:58 and SEQ ID NO:385 encode the polypeptides of SEQ ID NO:227 and SEQ ID NO:502, respectively, with amino acid sequence
MAAAAVPSLLLSLPPHQGLTFSNKIQPFGAQGV LHPEPGLRDWLLPTCSRQLRVALPEKGS
EGSLCQTQLPATPCFLPSNTVRT. It will be appreciated that all characteristics and uses of the
15 polynucleotides of SEQ ID NOs:58 and 385 and polypeptides of SEQ ID NO:227 and 502,
described throughout the present application also pertain to the human cDNA of clone 166601, and
the polypeptides encoded thereby

- The polynucleotides of SEQ ID NO:58 and 385 and polypeptides of SEQ ID NO:227 and 502 encode a transcriptional regulatory protein. Preferred polynucleotides of the invention include
20 the nucleic acid sequences comprising Clone 166601, the polynucleotides comprising SEQ ID
NO:58 and the polynucleotides comprising SEQ ID NO:385. Preferred polypeptides of the
invention include the amino acid sequences derived from the nucleic acid sequence comprising
Clone 166601, the polypeptides comprising the amino acid sequences of SEQ ID NO:227 and the
polypeptides comprising the amino acid sequences of SEQ ID NO:502.

- 25 In an embodiment of the invention, preferred polypeptides include the portion comprising
the site of protein kinase C phosphorylation or the amino acid sequences comprising SNK or TVR
of SEQ ID NO:227 and 502.

- In another embodiment, preferred polypeptides of the invention include the portion of the
amino acid sequence comprising sites of myristylation or the amino acids comprising the sequence
30 GLTFSN or GSEGSL of SEQ ID NO:227 or 502.

Proteins of SEQ ID NO:268 (internal designation Clone ID:211056) and related protein of SEQ ID
NO:530.

- The polynucleotides of SEQ ID NO:99 and SEQ ID NO:416 and polypeptides of SEQ ID
35 NO:268 and SEQ ID NO:530, respectively, encode a novel human tryptophan hydroxylase,
including the amino acid sequence hereafter referred to as nhTOH. Tryptophan is taken up by
active transport into the neurons where it is hydroxylated to 5-hydroxytryptophan (5HTP). The
latter is then decarboxylated to serotonin, a neurotransmitter involved in central nervous disorders,
especially mood disorders, sleep disorders, and eating disorders. Activity of the polypeptide of the

invention increases production of serotonin levels and increase the metabolism of tryptophan. Thus polypeptides of the invention are useful in the in vitro production of the serotonin and metabolism of tryptophan. As example, an expression vector containing the polynucleotides of SEQ ID NO:99 or SEQ ID NO:416 can be introduced into a cell line by methods known in the art such as by
5 calcium precipitation; tryptophan can be supplied in the media; and serotonin produced by the cells can be extracted by known methods.

The invention further relates to a method of screening for test compounds that bind hnTOH comprising the steps of contacting a hnTOH polypeptide with said test compound and detecting or measuring whether said test compound binds said hnTOH polypeptide. The invention further
10 relates to a method of screening for test compounds that activate hnTOH comprising the steps of contacting a hnTOH polypeptide with said test compound and detecting or measuring whether said test compound activates said hnTOH polypeptide, for example by measuring serotonin production or tryptophan depletion.

Another embodiment includes physiologically acceptable compositions of test compounds
15 found to increase serotonin production, referred to as activators, in a screen. Further embodiments include methods to use activators that have been identified in a screen or previously known in the art in the preparation of physiological acceptable formulations for use in in vivo. Further preferred are methods to use activators in a physiologically acceptable formulation in the treatment of CNS disorders in which tryptophan and serotonin levels are aberrant, particularly depression, anxiety
20 disorder, bipolar disorder, and eating disorders.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:99 and SEQ ID NO:416 and polypeptides of SEQ ID NO:268 and SEQ ID NO:530, described throughout the present application also pertain to the human cDNA of clone 211056, and the polypeptides encoded thereby.

25 Proteins of SEQ ID NO:190 (internal designation Clone ID:147648) and related protein of SEQ ID NO:474.

The polynucleotides of SEQ ID NO:21 and SEQ ID NO:357 and polypeptides of SEQ ID NO:190 and SEQ ID NO:474 encode a novel DNA binding polypeptide containing a leucine zipper
30 pattern multimerization domain, thereafter referred to as LZIP, also known as bZIP transcription factor basic domain signature (Hai et al., Genes Dev. 3:2083(1989)). An embodiment of the present invention includes the polynucleotides, polypeptides and fragments thereof comprising the sequences of SEQ ID NO:21, 357, 190, and 474 of the invention. Preferred polypeptides of the present invention are directed to the amino acid sequences which comprise the leucine zipper
35 domain selected from the following amino acids of SEQ ID NO:190 and 474 including LAAGAVTLGIGFFALASALWFL; PKGFFNYLTYFLAAGAVTLGIG; or FFALASALWFLICKRREIFQNS. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:21 and SEQ ID NO:357 and polypeptides of SEQ ID NO:190 and

SEQ ID NO:474, described throughout the present application also pertain to the human cDNA of clone 147648, and the polypeptides encoded thereby.

Leucine-zippers permit dimerization of various cytoplasmic hormone receptors and enzymes (Forman, et al., *Mol Endocrinol*, 3, 1610-1626 (1989)). Leucine zippers are also a
5 common feature of transcription factors, where they permit homo- or heterodimerization resulting in tight binding to DNA strands (for reviews, see Abel, et al., *Nature* 341, 24-25 (1989); Jones, et al., *Cell* 61, 9-11 (1990); Lamb, et al., *Trends in Biochemical Sciences* 16, 417-422 (1991)). Therefore, preferred polypeptides of the present invention are useful tools in several areas of biotechnology, especially in protein engineering, where their ability to mediate homo-dimerization or hetero-
10 dimerization has found several applications, including but not limited to immunochemistry, antibody generation, preparation of soluble oligomeric proteins, complementation assays. The utility of the present invention may be further confirmed by methods described, for example, by Bosslet et al (US patent 5,643,731) in which use of a pair of leucine zippers for in vitro diagnosis, in particular for the immunochemical detection and determination of an analyte in a biological liquid ;
15 by Tso et al (US patent 5,932,448) in which use of leucine zippers for producing bispecific antibody heterodimers ; by Conrad et al (US patent 5,965,712) , Ciardelli et al (US patent 5,837,816) , and Spriggs et al (WO9410308) in which methods of preparing soluble oligomeric proteins using leucine zippers have been described; and by Pelletier et al (WO9834120) in which methods to use leucine zipper forming sequences in protein fragment complementation assays to detect
20 biomolecular interactions has been described, all examples which disclosures are hereby incorporated by reference in their entireties.

The multimerization activity of the polypeptides of the present invention containing leucine zipper domains may be assayed using any of the assays known to those skilled in the art including circular dichroism spectrum and thermal melting analyses as described in US patent 5,942,433,
25 which disclosures are hereby incorporated by reference in their entirety. Alternatively, the leucine zipper motif in LZP could be used by those skilled in art as a "bait protein" in a well established yeast double hybridization system to identify its interacting protein partners in vivo from cDNA library derived from different tissues or cell types of a given organism. Alternatively, LZP or part thereof could be used by those skilled in art in mammalian cell transfection experiments. When
30 fused to a suitable peptide tag such as [His]₆ tag in a protein expression vector and introduced into culture cells, this expressed fusion protein can be immunoprecipitated with its potential interacting proteins by using anti-tag peptide antibody. This method could be chosen either to identify the associated partner or to confirm the results obtained by other methods such as those just mentioned.

In a preferred embodiment, the invention relates to compositions and methods of using the
35 LZP polynucleotides and polypeptides of SEQ ID NO: and SEQ ID NO: or fragment thereof for preparing soluble multimeric proteins, which consist in multimers of fusion proteins containing a leucine zipper fused to a protein of interest, using any technique known to those skilled in the art including those described in international patent WO9410308, which disclosure is hereby incorporated by reference in its entirety. In another preferred embodiment, LZP or derivative

thereof is used to produce bispecific antibody heterodimers as described in US patent 5,932,448, which disclosure is hereby incorporated by reference in its entirety. Briefly, leucine zippers capable of forming heterodimers are respectively linked to epitope binding components with different specificities. Bispecific antibodies are formed by pairwise association of the leucine zippers, forming an heterodimer which links two distinct epitope binding components. In still another preferred embodiment, LZP or part thereof or derivative thereof is used for detection and determination of an analyte in a biological liquid as described in US patent 5,643,731, which disclosure is hereby incorporated by reference in its entirety. Briefly, a first leucine zipper is immobilized on a solid support and the second leucine zipper is coupled to a specific binding partner for an analyte in a biological fluid. The two peptides are then brought into contact thereby immobilizing the binding partner on the solid phase. The biological sample is then contacted with the immobilized binding partner and the amount of analyte in the sample bound to the binding partner determined. In still another preferred embodiment, the LZP or part thereof may be used to synthesize novel nucleic acid binding proteins which are able to multimerize with proteins of interest, for example to inhibit and/or control cellular growth using any genetic engineering technique known to those skilled in the art including the ones described in the US patent 5,942,433, which disclosure is hereby incorporated by reference in its entirety.

In another embodiment, the invention relates to compositions and methods using the LZP or part thereof or derivative thereof in protein fragment complementation assays to detect biomolecular interactions in vivo and in vitro as described in international patent WO9834120, which disclosures is hereby incorporated by reference in its entirety. Such assays may be used to study the equilibrium and kinetic aspects of molecular interactions including protein-protein, protein-nucleic acid, protein-carbohydrate and protein-small molecule interactions, for screening cDNA libraries for binding to a target protein with unknown proteins or libraries of small organic molecules for biological activity.

Still, another object of the present invention relates to the use of the LZP or part thereof for identifying new leucine zipper domains using any techniques for detecting protein-protein interaction known to those skilled in the art. Among the traditional methods which may be employed are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns of cell lysates. Once isolated as a protein interacting with the LZP, such an intracellular protein can be identified (e.g. its amino acid sequence determined) and can, in turn, be used, in conjunction with standard techniques, to identify other proteins with which it interacts. The amino acid sequence thus obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for gene sequences encoding such intracellular proteins. Screening may be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and the screening are well-known. (See, e.g., Ausubel *et al.*, eds., *Current Protocols in Molecular Biology*, J.Wiley and Sons (New York, NY 1993) and PR Protocols: A Guide to Methods and Applications, 1990, Innis, M. et al., eds. Academic Press, Inc., New York).

Alternatively, methods may be employed which result in the simultaneous identification of genes which encode the intracellular proteins that can dimerize with the LZP or part thereof using any technique known to those skilled in the art. These methods include, for example, probing cDNA expression libraries, in a manner similar to the well known technique of antibody probing of
 5 lambda.gt11 libraries, using as a probe a labeled version of the LZP or part thereof, or fusion protein, e.g., the LZP or part thereof fused to a marker (e.g., an enzyme, fluor, luminescent protein, or dye), or an Ig-Fc domain (for technical details on screening of cDNA expression libraries, see Ausubel *et al, supra*). Alternatively, another method for the detection of protein interaction in vivo, the two-hybrid system, may be used.

10 Proteins of SEQ ID NO:318 (internal designation Clone ID:124608) and related protein of SEQ ID NO:556.

The polynucleotides of SEQ ID NO:149 and SEQ ID NO:442 and polypeptides of SEQ ID NO:318 and SEQ ID NO:556 encode an RNA-binding protein, hgRBP, which functions in RNA
 15 processing and protein expression. The preferred composition of SEQ ID NO:318 and 556 include MERPDKAALNALQPPEFRNESSLASTLKTLLFF TALMITVPIGLYFTTKSYIFEGALGMSNR DSYFYAAIVAVVAVHVVLALFVYVAWNEGSRQWREGKQD.

Further preferred polypeptides include those of SEQ ID NO:318 or SEQ ID NO:556 comprising an N-myristoylation site or the amino acid sequence at positions 43-48 or comprising
 20 the amino acid sequence GLYFTT which targets the protein to the membrane of the endoplasmic reticulum for function of hgRBP in translation of cellular mRNA into protein.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:149 and SEQ ID NO:442 and polypeptides of SEQ ID NO:318 and SEQ ID NO:556, described throughout the present application also pertain to the human cDNA 124608 of clone , and the
 25 polypeptides encoded thereby.

Protein of SEQ ID NO:337 (internal designation Clone ID:113448)

The polynucleotides of SEQ ID NO:168 and related SEQ ID NO:454 and polypeptides of SEQ ID NO:337 encode a novel human RNA-binding protein involved in RNA processing and
 30 protein expression which is related to Clone ID:183902 and Clone ID:635993.

It will be appreciated that all characteristics and uses of the The polynucleotides of SEQ ID NO:168 and related SEQ ID NO:454 and polypeptides of SEQ ID NO:337, described throughout the present application also pertain to the human cDNA of clone 113448, and the polypeptides encoded thereby.

35 Protein of SEQ ID NO:328 (internal designation Clone ID:183902)

Polynucleotides of SEQ ID NO:159 and related SEQ ID NO:450 and polypeptides of SEQ ID NO:328 encode a novel human RNA-binding protein involved in RNA processing and protein expression which is related to Clone ID:113448 and Clone ID:635993.

It will be appreciated that all characteristics and uses of the The polynucleotides of SEQ ID NO:159 and related SEQ ID NO:450 and polypeptides of SEQ ID NO:328, described throughout the present application also pertain to the human cDNA of clone 183902, and the polypeptides encoded thereby.

5

Protein of SEQ ID NO:329 (internal designation Clone ID:635993)

Polynucleotides of SEQ ID NO:160 and related SEQ ID NO:451 and polypeptides of SEQ ID NO:329 encode a novel human RNA-binding protein involved in RNA processing and protein expression which is related to Clone ID:183902 and Clone ID:113448. It will be appreciated that all characteristics and uses of the The polynucleotides of SEQ ID NO:160 and related SEQ ID NO:451 and polypeptides of SEQ ID NO:329, described throughout the present application also pertain to the human cDNA of clone 635993, and the polypeptides encoded thereby.

Many eukaryotic proteins that bind single-stranded RNA contain one or more copies of a putative RNA-binding domain of about 90 amino acids. This is known as the eukaryotic RNA-binding region, RNP-1 signature or RNA recognition motif (RRM) (Bandziulis et al. Genes Dev. 3:431 (1989); Swanson et al. Trends Biochem. Sci. 13: 86-91 (1988)). RRM's are found in a variety of RNA binding proteins, including heterogeneous nuclear ribonucleoproteins (hnRNPs), proteins implicated in regulation of alternative splicing, and protein components of small nuclear ribonucleoproteins (snRNPs). The polypeptides of SEQ ID NO:337, 328, and 329 encode novel human RNA binding protein, hereafter referred to as ghRBP which contains one copy of an RRM. Further characteristic of a protein which binds to nucleic acids, ghRBP contains a zinc finger motif comprising the amino acid sequence. Preferred polynucleotides of the invention include polynucleotides comprising the nucleic acids of SEQ ID NO:159, 160, 168, 450, 451, and 454. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:337, 328 and 329. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:159, 160, 168, 450, 451, and 454 described throughout the present application also pertain to the human cDNA of Clone ID:183902, Clone ID:635993 and Clone ID:113448.

Preferred amino acids of the invention are residues which comprise the RNA-binding domain or portion thereof. Preferred amino acid sequences are selected from the following set of sequences including

AFVRRXPWTAASSQLKEHFAQFGHVRRRCILPFDKETGFHRLGLGWVQFSSEGLRNALQQE
NHIIDGVKVQV;
SINQPVAFVRRXPWTAASSQLKEHFAQFGHVRRRCILPFDKETGFHRLGLGWVQFSSEGLRN
ALQQENHIID;
PWTAASSQLKEHFAQFGHVRRRCILPFDKETGFHRLGLGWVQFSSEGLRNALQQENHIIDGV
KVQVHTRRP. Further preferred are polypeptides of the invention include any fragment of SEQ ID NO: which binds to RNA.

An embodiment of the invention relates to methods of using the polypeptides of the invention to bind to RNA molecules in vitro by techniques that are known in the art. Preferred use

of the polypeptides of the invention includes extraction of RNA from biological samples, chemical reagents, cell homogenates and tissue homogenates. Further utility of the polypeptides of the present invention or part thereof may be further confirmed by binding methods described in Trifillis, et al., RNA 5(8): 1071-82 (1999) and U.S. Patent 6,107,029, which disclosures are hereby
 5 incorporated by reference in their entireties.

Proteins of SEQ ID NO:248 (internal designation Clone ID:199782), SEQ ID NO:249 (internal designation Clone ID:821212), SEQ ID NO:250 (internal designation Clone ID:202863) and related protein of SEQ ID NO:518.

10 The polynucleotides of SEQ ID NOs:79, 80, 81 and 401 and polypeptides of SEQ ID NOs:248, 249, 250 and 518 encode human RNA-associated polypeptides which act as splicing factors. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:79, 80, 81 and 401 and polypeptides of SEQ ID NOs:248, 249, 250 and 518, described throughout the present application also pertain to the human cDNA of clones 199782, 821212, and
 15 202863, and the polypeptides encoded thereby.

The translation of genetic information into protein depends on RNA and the first step in this process is the transcription of DNA into RNA while retaining all the genetic information encoded in DNA. The RNA transcript undergoes various processing steps which include splicing and polyadenylation. The mature RNA transcript is translated into protein by the ribosomal machinery.
 20 Nascent RNA transcripts are spliced in the nucleus by the spliceosomal complex which catalyzes the removal of introns and the rejoining of exons. At least 40 splicing factors have been identified and interaction of these factors are important in the conformational changes needed for the enzymatic removal of introns and religation of the exons. Both protein and RNA components are involved in the spliceosome assembly and the splicing reaction. There are 2 distinct catalytic steps
 25 involved in the RNA splicing reaction with distinct proteins and RNA species. Alternative splicing factors include developmentally regulated proteins that play key roles in developmental processes such as pattern formation and sex determination, respectively (Hodgkin, J. et al. (1994) Development 120:3681-3689). Alternate splicing is also involved in the tissue specific expression of isoforms of proteins, including structural proteins and enzymes.

30 An embodiment of the present invention relates to compositions of the polynucleotides of SEQ ID NO:79, 80, 81 and 401 and polypeptides of SEQ ID NO:248, 249, 250 and 518. Preferred amino acids of the invention comprise the zinc finger region or fragment thereof and are selected from the following sequences of amino acids from SEQ ID NO:248, 249, 250 and 518 including GACENCGAMTHKKKDCFE; NSITKYRKGACENCGAM; or THKKKDCFERPRRVGAKF.

35 The polypeptides of SEQ ID NO:248, 249, 250 and 518 are involved in the spliceosome complex and have function in the processing of RNA processing. Alternatively, the polypeptides of the present invention are involved in RNA processing and thus involved in protein expression. A preferred embodiment of the invention relates to a method of using the polynucleotides of polynucleotides of SEQ ID NO:79, 80, 81 and 401 in vitro. A preferred method of use relates to

introduction of said polypeptides or fragments thereof into cells by techniques known in the art such as transfection or microinjection. Further preferred are methods to use the polynucleotides of the invention to alter protein expression in the given cell. Alternately, polypeptides of the present invention can be used in combination with reagents known in the art to alter protein expression in
5 cell free expression systems, mammalian expression systems, insect expression systems, or bacterial expression systems. Furthermore, methods to use the polynucleotides or polypeptides the present invention to increase or decrease protein expression is preferred. The utility of the polypeptides of the invention or part thereof may be further confirmed using methods described in US Patent 6020164 and Chua and Reed, Gene Devel 13:841-850 (1999), which disclosures are hereby
10 incorporated by reference in their entireties.

In another embodiment, methods to screen for inhibitors and activators of the polypeptides of the invention are preferred. In another embodiment, molecules or compounds which are identified in such a screen are further preferred. Further preferred are compounds which activate or inhibit activity of the polypeptides of the current invention. Activity of the the polypeptides of the
15 invention is modified by phosphorylation at cAMP and cGMP dependent phosphorylation sites (including 3-6:55-58; 107-110) and casein kinase II phosphorylation sites (including 33-36:58-61:126-129). Preferred activators include but are not limited to compounds which promote accumulation of intracellular cAMP and cGMP. Further preferred activators include those compounds which activate casein kinase II. Preferred inhibitors are those compounds which inhibit
20 intracellular cAMP and cGMP accumulation, or those compounds which promote cAMP and cGMP degradation. Further preferred inhibitors include compounds which promote the deactivation of casein kinase II. Furthermore, inhibitors and activators of the polypeptides of the present invention include compounds known in the art as well as compounds to be identified by the method of screening. Furthermore, compounds that inhibit or activate the activity of the polypeptides of the
25 present invention by means other than phosphorylation or dephosphorylation are also preferred.

In another embodiment, a method for the use of the polynucleotides or polypeptides of the present invention in the treatment, prevention, attenuation or diagnosis of disorders of RNA processing or protein processing are preferred. Such disorders are selected from a group which includes but is not limited to cancers such as adenocarcinoma, leukemia, sarcoma, teratocarcinoma,
30 and any disorder associated with cell growth and differentiation, embryogenesis, and morphogenesis involving any tissue, organ, or system, e.g., the brain, adrenal gland, or reproductive system.

Proteins of SEQ ID NO:256 (internal designation Clone ID:822794), SEQ ID NO:257 (internal designation Clone ID:337572) and related protein of SEQ ID NO:521.
35

The polynucleotides of SEQ ID NO:87, 88 and 407 and polypeptides of SEQ ID NO:256, 257 and 521 encode human nuclear polypeptides which interact with transcription factors of the Signal Transducers and Activators of Transcription (STAT) family of proteins involved in the regulation of cell division. It will be appreciated that all characteristics and uses of the

polynucleotides of SEQ ID NO:87, 88 and 407 and polypeptides of SEQ ID NO:256, 257 and 521, described throughout the present application also pertain to the human cDNA of clones 822794 and 337572, and the polypeptides encoded thereby.

STATs are pleiotropic transcription factors which mediate cytokine-stimulated gene
5 expression in multiple cell populations (Levy, Cytokine Growth Factor Rev., 8:81 (1997)). All STAT proteins contain a DNA binding domain, a Src homology 2 (SH2) domain, and a transactivation domain necessary for transcriptional activation of target gene expression. Janus kinases (JAK), including JAK1, JAK2, Tyk, and JAK3, are cytoplasmic protein tyrosine kinases (PTKs) which play pivotal roles in initiation of cytokine-triggered signaling events by activating the
10 cytoplasmic latent forms of STAT proteins via tyrosine phosphorylation on a specific tyrosine residue near the SH2 domain (Ihle et al., Trends Genet., 11: 69 (1995); Darnell, Science 277(5332):1630 (1997); Johnston et al., Nature, 370: 1513 (1994)). Tyrosine phosphorylated STAT proteins dimerize through specific reciprocal SH2-phosphotyrosine interactions and translocate from the cytoplasm to the nucleus where they stimulate the transcription of specific target genes by
15 binding to response elements in their promoters (Leonard, Nature Medicine, 2: 968 (1996); Zhong et al., PNAS USA, 91:4806 (1994) Darnell, Science, 277:1630 (1997)).

In an embodiment of the present invention, compositions of the polynucleotides and polypeptides or fragments thereof SEQ ID NO:87, 88 and 407 and SEQ ID NO:256, 257 and 521, respectively are included. Further preferred are polypeptides of the present invention which interact
20 with activated STAT 3, but may also interact with STAT1, STAT2 or other STAT homologues. Preferred polypeptides of the invention act to inhibit or decrease the activity of STATs. Further preferred amino acids of SEQ ID NO:256, 257 and 521 include the SAP domain VSSFRVSELQVLLGFAGRNKSGRKHDLMLRALHLL. Activation of cytokine receptors by their cognate ligands activate JAKs which in turn, activate STATs. Therefore cytokines and other
25 hormones which signal through cytokine-like receptors may be modulated by polypeptides or polynucleotides of the present invention. Cytokines and other hormones which can thus be modulated by the present invention include but are not limited to interferons, interleukins, prolactin, and growth hormone. The utility of the polypeptides of the present invention or part thereof may be further confirmed using the methods described in WIPO Publication WO9928465 which disclosure
30 is hereby incorporated by reference in its entirety.

The polypeptides of this invention can be used in a method of inhibiting the activity of STAT proteins in a cell in vitro, the method comprising introducing a nucleic acid into the cell, wherein the nucleic acid comprises a nucleotide sequence encoding the amino acid sequence of
35 SEQ ID NO:256, 257 and 521 or the amino acid sequence of SEQ ID NO:256, 257 and 521 with one or more conservative amino acid alterations, and wherein the nucleic acid expresses the amino acid sequence in an amount and for a time sufficient for the amino acid sequence to specifically bind to STAT proteins and to decrease STAT activity, thereby decreasing STAT activity in the cell.

Suitable compositions of polypeptides or polynucleotides of the present invention are useful as a method of treatment of pathologies such as diseases, syndromes, or other undesirable

conditions resulting from defects in cell cycle progression. Such cell cycle defects may result from defects in the regulation of activated STAT or an upstream factor such as activated JNKs or activated cytokine receptors. Alternatively, polypeptides or polynucleotides of the present invention may be used in a method of treating pathologies resulting from defects in cell cycle progression due to defects in a step "downstream" of STAT regulation of cell cycle progression. In preferred embodiments, agonists of polypeptides or polynucleotides of the present invention are useful in the treatment of pathologies such as but not limited to hyperproliferative diseases such as cancer (e.g., leukemia, lymphoma, breast cancer, colon cancer, prostate cancer, Wilms' tumor), coronary artery disease, pulmonary vascular obstructive disease, either primary or as a feature of Eisenmenger's syndrome, and other disorders of abnormal cellular proliferation. Cells to be treated include but are not limited to hyperproliferative cells, cancer cells, vascular smooth muscle cells, endothelial cells, and gametes.

In some embodiments of the invention, antagonists of the polypeptides or polynucleotides of the present invention are used to stimulate, promote, or facilitate progression through the cell cycle, such as in the cellular regeneration of terminally differentiated cardiac myocytes or tissues, e.g., striated muscle myocytes. For example, this could allow restoration of damaged myocardium after cardiac injury, myocardial infarction, myocarditis, cardiomyopathy, trauma, as a consequence of cardiac surgery, etc., or repletion of striated muscle exhausted by muscular dystrophy.

In further embodiments, expression of the polypeptides encoded by the nucleic acids is expected to prevent, ameliorate, or lessen the cell cycle defect of the host cell, or to restore normal cell cycle progression of the host cell. Whether provided via nucleic acid or polypeptides delivered directly to cells, the therapeutic formulations of the invention can also be used as adjuncts to other forms of therapy, including but not limited to chemotherapy, and radiation therapy.

25 Protein of SEQ ID NO:330 (internal designation Clone ID:398703) and related protein of SEQ ID NO:330.

The polynucleotides of SEQ ID NO:161 and SEQ ID NO:452 and polypeptides of SEQ ID NO:330 encode a novel human deubiquitinating enzyme (GNP:AF017306). Deubiquitinating enzymes serve a number of functions (Hochstrasser *Cur Opin Cell Biol* 4:1024 (1992); Rose, In: Ubiquitin, Plenum Press, New York (1988)). First, ubiquitin must be cleaved from a set of biosynthetic precursors, which occur either as a series of ubiquitin monomers in head-to-tail linkage or as fusions to certain ribosomal proteins (Finley & Chau, *Annu Rev Cell Biol* 7, 25-69 (1991)). Secondly, ubiquitin must be recycled from intracellular conjugates, both to maintain adequate pools of free ubiquitin and, in principle at least, to reverse the modification of inappropriately targeted proteins. Finally, deubiquitinating reactions may be integral to the degradation of ubiquitinated proteins by the 26S proteasome, a complex ATP-dependent enzyme whose exact composition and range of activities remain poorly characterized (Hershko & Ciechanover, *Annu Rev Biochem* 61, 761-807 (1992); Hadari et al., *J Biol Chem* 267, 719-727 (1992); Murakami et al., *Nature* 360, 597-9 (1992); Rechsteiner, *J. Biol. Chem.* 268, 6065-6068 (1993)).

An embodiment of the invention includes preferred polypeptides with ubiquitin-specific protease activity with a novel N-terminus of Clone ID:398703 comprising the amino acid sequence MCTTSLPCPIIMEPWGLATTKAAYVLFYQRRDDEFYKTPSLSSSGSSDGGTRPSSSQQGGFD DEACSMMDTN encoded by

5 ATGTGTACGACCTCATTGCCGTGTCCAATCATTATGGAGCCATGGGGGTTGGCCACTAC
TAAAGCAGCTTATGTGCTATTTTACCAACGTCGAGATGATGAATTTTATAAGACACCTT
CACTTAGCAGTTCTGGTTCCTCTGATGGAGGGACACGACCAAGCAGCTCTCAGCAGGG
CTTTGGGGATGATGAGGCTTGACGATGGACACCAACTAA of SEQ ID NO:161.

The preferred polypeptides of the invention are those which prevent or reverse
10 ubiquitination of cellular proteins in vitro or in vivo. Further preferred are polypeptides of the invention which prevent or reverse ubiquitination of extracellular proteins in vitro or in vivo.

The polynucleotides of SEQ ID NO:161 and SEQ ID NO:452 encode polypeptides of SEQ ID NO:330 which contain protein domains or motifs including but not limited to a Protein kinase C phosphorylation site comprising the amino acid fragment SAR, and an N-myristylation site
15 comprising amino acid fragment GLNMSE. Further preferred amino acids of SEQ ID NO : 330 include the Ubiquitin carboxyl-terminal hydrolases family 2 signature or amino acid sequence YDLIAVSNHYGAMGVGHY.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:161 and SEQ ID NO:452 and polypeptides of SEQ ID NO:330, described throughout the
20 present application also pertain to the human cDNA of clone 398703, and the polypeptides encoded thereby.

The utility of the polynucleotides and polypeptides of the present invention or part thereof may be further confirmed using methods which assess activity or function of deubiquitinating enzymes described in United States Patents 5391490 and 5565352 which disclosures are hereby
25 incorporated by reference in their entireties.

Proteins of SEQ ID NO:277 (internal designation Clone ID:653966) and related protein of SEQ ID NO:535.

The polynucleotides of SEQ ID NO:108 and SEQ ID NO:421 and polypeptides of SEQ ID
30 NO:277 and SEQ ID NO:535 encode human liver fatty acid binding protein (L-FABP) comprising the amino acid sequence of SEQ ID NO:277 and 535. The amino acid sequence of SEQ ID NO:277 and 535 are the same as human L-FABP (Genbank accession GNP:M10617; Lowe et al., JBC 260:3413-17 (1985)) and homologous to human FABP (Genbank accession GNP:M10050). The polypeptides of the present invention belong to the FABP/P2/CRBP/CRABP family of transporters
35 and functionally binds to free fatty acids and derivatives thereof. L-FABP is normally expressed in the cytoplasm of hepatocytes, but preferred embodiments include use of the polypeptides of the present invention as extracellular polypeptides. Further preferred embodiments include use of the polypeptides of the present invention as serum or plasma polypeptides. Further preferred

embodiments use polypeptides of the invention in vitro. Still further preferred embodiments include use of the polypeptides of the present invention in vivo.

Preferred amino acids of the invention include the lipocalin domain, from 2 to 127 or polypeptides comprising the amino acid sequence

5 SFSGKYQLQSQENFEAFMKAIGLPEELIQKGKDIKGVSEIVQNGKHFKFTITAGSKVIQNEFT
VGEECELETMTGEKVKT VVQLEGDNKLVTTFKNIKSVTELNGDIITNTMTLGDIVFKRISKR

I. The polypeptides of SEQ ID NO:277 and SEQ ID NO:535 contain a cytosolic fatty-acid binding protein signature comprising the amino acid sequence GKYQLQSQENFEAFMKAI which functions in the polypeptides ability to bind small hydrophobic molecules, such as lipids, steroid hormones, and retinoids. Preferred amino acids of SEQ ID NO:277 and SEQ ID NO:535 include 10 GKYQLQSQENFEAFMKAI, MSFSGKYQLQSQENFEAF, and LQSQENFEAFMKAIGLPE.

Phosphorylation status modulates the activity of L-FABP. Preferred polypeptides of the invention include the amino acids sequence comprising the sites of cAMP- and cGMP-dependent protein kinase phosphorylation including residues of SEQ ID NO:277 and 535 comprising the 15 sequence KRIS.

Further preferred polypeptides of the invention include the amino acid sequence comprising the sites of Protein kinase C phosphorylation including residues at positions 4 to 6, 94 to 96, and 124 to 126 of SEQ ID NO:277 and 535. Still further preferred polypeptides of SEQ ID NO:277 and 535 include the amino acid sequence comprising SGK, TFK, and SKR.

20 Further preferred are polypeptides of SEQ ID NO: 277 and 535 include the amino acid sequence comprising a Casein kinase II phosphorylation sites. Preferred amino acids of SEQ ID NO: 277 and 535 include positions 64 to 67, 100 to 103, and 114 to 117. Further preferred amino acids comprise the sequences TVGE, SVTE, and TLGD.

A preferred polypeptide of SEQ ID NO: 277 and 535 is one in which the amino acid 25 asparagine (Asn) is located at residue 105, further referred to as the N-isoform. Further preferred is the polypeptide of SEQ ID NO: 277 and 535 in which the amino acid aspartate (Asp) is located at residue 105 further referred to as the D-isoform. The rat homologue of the human D-isoform of the present invention was shown to have a greater affinity to lysophospholipids, prostaglandins, retinoids, bilirubin and bile salts compared to the rat homologue of the human N-isoform of the 30 present invention by methods described by DiPietro and Santome, Biochim Biophys Acta 1478 :186-200 (2000) which disclosure is hereby incorporated by reference in its entirety. The rat homologues share only 82% identity with the of the human D- and N-isoforms, therefore it is not predictable to find that the human D-isoform has equal or greater affinity to lysophospholipids, prostaglandins, retinoids, bilirubin, bile salts and fatty acid compared to the human N-isoform.

35 Further preferred polypeptides of the present invention include the D-isoform polypeptide and fragments thereof which have an equal or at least 10%, 20%, 30%, 40%, 50%, 60% or 75% greater affinity for fatty acids, and lipophilic compounds selected from a group including but not limited to lysophospholipids, prostaglandins, retinoids, bilirubin, bile salts, steroid hormones (such as testosterone and estradiol), and cholesterol compared to the N-isoform.

Another embodiment of the invention includes polynucleotides or polypeptides of the invention or fragments thereof which bind lipophilic compounds selected from a group including but not limited to free fatty acids, lysophospholipids, prostaglandins, retinoids, bilirubin, bile salts, steroid hormones (such as testosterone and estradiol), and cholesterol in serum or plasma. Further preferred are polypeptides of the invention which bind lipophilic compounds in serum or plasma separated from whole blood in a process of purifying serum or plasma for use in vitro or in vivo. Further preferred are polypeptides of the invention which bind lipophilic compounds in serum or plasma in vivo.

In a further embodiment, polynucleotides or polypeptides of the invention or fragments thereof, in physiological appropriate formulations, are useful in the prevention, treatment or attenuation of conditions in which lipophilic compounds are elevated in the serum of mammals, preferably humans. Such conditions are selected from a group which include but are not limited to obesity, hyperlipidemia, hypercholesterolemia, hypertriglyceridemia, diabetes type I (IDDM) diabetes type II (NIDDM), atherosclerosis, and hypertension.

Mammary-derived growth inhibitor (MDGI) and heart-fatty acid binding protein (*FABP*), which belong to the *FABP* family, specifically inhibit growth of normal mouse mammary epithelial cells (MEC) and promote morphological differentiation, stimulates its own expression and promotes milk protein synthesis (US Patent 5977309, 24 March 1995). In further preferred embodiments, polypeptides of the invention include those which locally signal growth cessation and stimulate differentiation of the developing epithelium. Further preferred polypeptides of the invention suppress the mitogenic effects of EGF family members, and inhibit c-fos, c-myc and c-ras expression.

In a further aspect of the present invention, there is provided a method for producing such polypeptide by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a human fatty acid binding polypeptides or polynucleotides of the invention acid under conditions promoting expression of said protein and subsequent recovery of said protein.

In a further embodiment of the present invention, there is provided a method for utilizing such polypeptides, or polynucleotides of the invention for therapeutic purposes, for example, as a cell growth inhibitor and as to cause differentiation stimulatory activity on various responsive types of tissues and cells in vitro. Further preferred are methods for use of polypeptides, or polynucleotides of the invention, in appropriate physiological form, for therapeutic purposes for to inhibit cell proliferation or to induce cell differentiation in mammals, preferably humans.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:108 and SEQ ID NO:421 and polypeptides of SEQ ID NO:277 and SEQ ID NO:535, described throughout the present application also pertain to the human cDNA of clone 653966, and the polypeptides encoded thereby.

Proteins of SEQ ID NO:313 (internal designation Clone ID:633418), SEQ ID NO:314 (internal designation Clone ID:422878) and related protein of SEQ ID NO:552.

The polynucleotides of SEQ ID NO:144, 145 and 438 and polypeptides of SEQ ID NO:313, 314 and 552 encode a cleavage stimulation factor important in mRNA processing and protein expression. Protein kinase C phosphorylation increases activity of said polypeptides and preferred amino acids include SEK and SGR. Further, sites of tyrosine kinase phosphorylation increase activity of said polypeptides and preferred amino acids of SEQ ID NO:314, 552 include KKLEENPY.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:144, 145 and 438 and polypeptides of SEQ ID NO:313, 314 and 552, described throughout the present application also pertain to the human cDNA of clones 633418 and 422878, and the polypeptides encoded thereby.

Proteins of SEQ ID NO:219 (internal designation Clone ID:589848), SEQ ID NO:220 (internal designation Clone ID:211883), SEQ ID NO:221 (internal designation Clone ID:642603), SEQ ID NO:222 (internal designation Clone ID:193316), and related protein of SEQ ID NO:497.

Polynucleotides of SEQ ID NO:50, 51, 52, 53, 380 and polypeptides of SEQ ID NO:219, 220, 221 and 497 encode RNA associated proteins with a ribosomal L34 domain comprising the amino acid sequence NEYQPSNIKRKNKHGWVRLXTPAGXXXILRRMLKGRKSLSH or NEYQPSNIKRKNKHGWVRLXTPAGVQVILRRMLKGRKSLSH. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:50, 51, 52, 53, 380 and polypeptides of SEQ ID NOs:219, 220, 221 and 497, described throughout the present application also pertain to the human cDNA of clones 589848, 211883, 642603 and 193316, and the polypeptides encoded thereby.

Proteins of SEQ ID NO:302 (internal designation Clone ID:1000891255) and related protein of SEQ ID NO:543.

Polynucleotides of SEQ ID NO:133 and 429 and polypeptides of SEQ ID NO:302 and 543 encode human ribosomal protein, hRIBPRT. An embodiment of the invention includes the compositions of the polypeptides of SEQ ID NO:302 and 543, comprising the amino acid sequence MVAACKTKKSLESIKSRLQLVMKSGKYVLGYKQTLKMIRQGKAKLVILANNCPALRKSEIEYYAMLAKTGVHHYSGNNIELGTACGKYRVCTLAIDPXDSXIIRSMPEQTGEK, and the polynucleotides of SEQ ID NO:133 and 429, respectively, which encode human ribosomal protein, hRIBPRT. The polypeptides of the invention contain the ribosomal protein L30e/L7Ae/S12e/Gadd4 signature (Koonin EV, J Mol Med 75:236-238 (1997) and Nakanishi et al., Gene 35:289-96 (1985)).

Preferred polypeptides of the invention include the amino acid sequence comprising KSLESIKSRLQLVMKSGKYVLGYKQTLKMIRQGKAKLVILANNCPALRKSEIEYYAMLAK

TGVHHYSGNNIELGTACGKYRVCTLAIDPXDSXIIR ;

KSLESIKRLQLVMKSGKYVLGYKQTLKMIRQGKAKLVILANNCPALRK ; and

SEIEYYAMLAKTGVHHYSGNNIELGTACGKYRVCTLAIDPXDSXIIR. Further preferred amino acids of the invention include sites of PKC phosphorylation, comprising the amino acid sequences of SEQ ID NO:302 and 543 including TKK (positions 7-13); SIK (positions 13-15); SGK (positions 24-26); and TLK (positions 34-36). Further preferred amino acids of the invention include sites of Casein Kinase II phosphorylation, comprising the amino acid sequences SEIE (positions 58-61) and SMPE (positions 107-110).

In another embodiment, the proteins of SEQ ID NO:302 and 543 can be used to bind to nucleic acids, preferably RNA, alone or in combination with other substances. For example, the proteins of the invention or part thereof can be added to a sample containing RNAs in optimum conditions for binding, and allowed to bind to RNAs. In a preferred such embodiment, the proteins of the invention or part thereof may be used to purify mRNAs, for example to specifically isolate RNA, e.g. from a specific cell type or from cells grown under particular conditions. Such RNAs could then be reverse transcribed and cloned, could be analyzed for relative expression analyses, etc. In addition, such methods may be used to specifically remove RNA from a sample, for example during the purification of DNA. To carry out any of these methods, the proteins of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other RNA binding proteins, to form an affinity chromatography column. A sample containing a mixture of nucleic acids to purify is then run through the column. Immobilizing the proteins of the invention or part thereof on a support is particularly advantageous for embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of RNAs from the batch of resin-coupled protein after binding, and allows subsequent re-use of the protein. Immobilization of the proteins of the invention or part thereof can be accomplished, for example, by inserting any matrix binding domain in the protein according to methods known to those skilled in the art. The resulting fusion product including the proteins of the invention or part thereof is then covalently, or by any other means, bound to a protein, carbohydrate or matrix (such as gold, "Sephadex" particles, polymeric surfaces).

Another embodiment of the present invention relates to methods and compositions using the proteins of the invention, or part thereof, to associate specific mRNAs to the inner face of lipidic bilayers of liposomes in order to further introduce these mRNAs into the cytoplasm of eukaryotic cells. Preferably, specific mRNAs are first associated with the protein of the invention and the RNA/protein complex formed in that way is then mixed with liposomes according to methods known to those skilled in the art. These liposomes are added to an *in vitro* culture of eukaryotic cells. In vivo, such a method might treat and/or prevent disorders linked to dysregulation of gene transcription such as cancer and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration.

A decrease in ribosome function results in a significant inhibition of cell growth. Therefore, in another embodiment, the present proteins and nucleic acids can be used to modulate

the rate of cell growth in vitro or in vivo. Accordingly, compounds that inhibits the expression or function of the proteins of the invention can be used to inhibit the growth rate of cells, and can thus be used, e.g. in the treatment or prevention of diseases or conditions associated with excessive cell growth, such as cancer or inflammatory conditions. Such compounds include, but are not limited to, antibodies, antisense molecules, dominant negative forms of the proteins, and any heterologous compounds that inhibit the expression or the activity of the proteins.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:133 and 429 and polypeptides of SEQ ID NO:302 and 543, described throughout the present application also pertain to the human cDNA of clone 1000891255, and the polypeptides encoded thereby.

Proteins of SEQ ID NO:271 (Internal designation Clone ID:493328), related clones 153261, 152042, 599054, and 650872 and related protein of SEQ ID NO:533.

The polynucleotides of SEQ ID Nos:102, 103, 104, 105, and 106 and polypeptides of SEQ ID Nos:271, 272, 273, 274, 275 and 533 encode for the HUMAN GENSET BINDING PROTEIN or HGBP-1. Preferred polypeptides comprise the amino acid sequence MKVKIKCWNGVATWLWVANDENCGICRMAFNGCCPDCKVPGDDCPLVWGQCSHCFHM HCILKWLHAQQVQQHCPMCRQEWFKE. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID Nos:102, 103, 104, 105, and 106 and polypeptides of SEQ ID Nos:271, 272, 273, 274, 275 and 533, described throughout the present application also pertain to the human cDNA of clones 493328, 153261, 152042, 599054, and 650872, and the polypeptides encoded thereby.

The protein of SEQ ID NO: 271 encoded by the extended cDNA SEQ ID NO:102 is the same as a hepatocellular carcinoma associated ring finger protein (EMBL AF247565) and Genset protein in WO0100806 (Genpep accession AX061622) with homology to an anaphase-promoting complex (APC) subunit from Drosophila (Embl accession number AJ251510). In addition, HGBP-1 exhibits the pfam PHD zinc finger signature from positions 33 to 79.

Zinc binding domains which contain a C₃HC₄ sequence motif are known as RING domains (Lovering, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90:2112-2116). Zinc finger domains are found in numerous zinc binding proteins which are involved in protein-protein and protein-nucleic acid interactions. They are independently folded zinc-containing mini-domains which are used in a modular repeating fashion to achieve sequence-specific recognition of DNA (Klug 1993 Gene 135, 83-92). Such zinc binding proteins are commonly involved in the regulation of gene expression, and usually serve as transcription factors, either by directly affecting transcription or recruiting co-activators or co-repressors (see US patents 5,866,325; 6,013,453 and 5,861,495). PHD fingers are C₄HC₃ zinc fingers spanning approximately 50-80 residues and distinct from RING fingers or LIM domains. They are thought to be mostly DNA or RNA binding domain but may also be involved in protein-protein interactions (for a review see Aasland et al, Trends Biochem Sci 20:56-59 (1995)).

HGBP-1 or part thereof is a zinc binding protein, which is able to bind nucleic acids, more preferably a transcription factor. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 271 from positions 33 to 79. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 271 having any of the biological
5 activity described herein. The nucleic acid binding activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in US patent 6,013,453.

The invention relates to methods and compositions using the protein of the invention or part thereof to bind to nucleic acids, preferably DNA, alone or in combination with other substances.
10 For example, the protein of the invention or part thereof is added to a sample containing nucleic acid in conditions allowing binding, and allowed to bind to nucleic acids. In a preferred embodiment, the protein of the invention or part thereof may be used to purify nucleic acids such as restriction fragments.

In another preferred embodiment, HGBP polypeptides or parts thereof may be used to
15 visualize nucleic acids when the polypeptide is linked to an appropriate fusion partner, or is detected by probing with an antibody. Thus, HGBP polypeptides can be used to diagnose...

Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other DNA binding proteins, using techniques well known in the art, to form an affinity chromatography column. A sample containing
20 nucleic acids to purify is run through the column. Immobilizing the protein of the invention or part thereof on a support advantageous is particularly for those embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of the protein from the batch of product and subsequent reuse of the protein. Immobilization of the protein of the invention or part thereof can be accomplished, for example, by inserting a cellulose-binding domain
25 in the protein. One of skill in the art will understand that other methods of immobilization could also be used and are described in the available literature.

In another embodiment, the present invention relates to compositions and methods using the protein of the invention or part thereof, especially the zinc binding domain, to alter the expression of genes of interest in a target cells. Such genes of interest may be disease related genes, such as
30 oncogenes or exogenous genes from pathogens, such as bacteria or viruses using any techniques known to those skilled in the art including those described in US patents 5,861,495; 5,866,325 and 6,013,453.

In a further embodiment, the protein of the invention or part thereof may be used to diagnose, treat and/or prevent disorders linked to dysregulation of gene transcription such as cancer
35 and other disorders relating to abnormal cellular differentiation, proliferation, or degeneration, including hyperaldosteronism, hypocortisolism (Addison's disease), hyperthyroidism (Grave's disease), hypothyroidism, colorectal polyps, gastritis, gastric and duodenal ulcers, ulcerative colitis, and Crohn's disease. The invention relates to methods of diagnosing, treating and/or preventing disorders described herein, comprising delivering to a patient, or causing to be present therein, a

zinc finger polypeptide which inhibits the expression of a gene enabling the cells to divide. The target could be, for example an oncogene or a normal gene, which is overexpressed in the cancer cells.

5 ISPG Iron-sulfur cluster protein (Clone ID:1000872335)

The polynucleotides of SEQ ID NOs:43 and 374 encodes the amino acids sequence of SEQ ID NOs:212 and 491 respectively, an iron-sulfur protein which mediates electron transfer in metabolic reactions, also referred to as ISPG. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs 43 and 374 and polypeptides of SEQ ID NOs:212 and 491, described throughout the present application also pertain to the human cDNA of clone 1000872335, and the polypeptides encoded thereby.

ISPG is an iron-sulfur protein that belongs to the broad family of the 2Fe-2S-type ferredoxins. The 2Fe-2S-type ferredoxins are proteins or domains of around one hundred amino acid residues that bind a single 2Fe-2S iron-sulfur cluster and are found in plants, animals and bacteria. Iron-sulfur cluster proteins are well known classes of proteins and are recognized as ideal devices for accepting, donating, storing and shifting electrons. It will be appreciated by the skilled artisan that structural aspect of iron-sulfur proteins have been studied extensively (reviewed in Beinert et al, Nature 1 Aug 1997, 277:653-659), allowing modifications of the ISPG proteins to fine tune its properties for desired uses while retaining its biological function in mediating electron transfer and possible protein stabilization and iron or sulfide storage functions. The ISPG protein of SEQ ID NO 491 comprises a glycosaminoglycan attachment site at amino acid position 34 (SGSG); protein kinase C phosphorylation sites at amino acid positions 14 (SAR), 44 (TTR), and 86 (SGR); N-myristoylation sites at amino acid positions 11 (GGVSAR), 24 (GTXWNR), 31 (GGTSGS), 39 (GVALGT), 106 (GACEAS), a cytochrome c family heme-binding site at amino acid positions 114-119 and an iron-sulfur binding region signature at amino acid positions 108-118.

In view of its role in electron transfer reactions, ISPG is thought to be involved in a wide variety of metabolic reactions and disorders, and may be useful in the treatment of disorders of metabolism such as obesity, in the detection of toxic compounds, in prediction, diagnosis or treatment of conditions or traits related to drug metabolism or in treatments related to the synthesis of eg. steroid hormones.

In a preferred example, overexpression or administration of the ISPG protein may be used as a therapeutic treatment for obesity by accelerating the metabolic rate of a subject in need of treatment. There is accumulating evidence to support the hypothesis that a low-energy-output phenotype is at high risk of weight gain and obesity, irrespective of whether this is owing to a low resting metabolic rate and/or physical inactivity. The low-energy-output phenotype is associated with impaired appetite control, which is improved if energy output is increased, serving as the background for pharmacologic stimulation of energy expenditure as a tool to improve the results of obesity management. The ISPG protein and agonists or stimulators thereof may serve as a means to

increase electron transfer and hence the metabolic rate of an individual in a similar goal as commonly cited targets such as leptin receptors, the sympathetic nervous system and its peripheral beta-adrenoceptors, selective thyroid hormone derivatives, and stimulation of the mitochondrial uncoupling proteins.

5 In addition, iron-sulfur proteins such as ISPG are generally recognized as being capable of several functions that are not of an oxidoreductive nature such as the binding and activation of substrates at the unique iron site (in the catalytic function of aconitase and related enzymes), and apparently stabilizing radicals in reactions occurring by a free-radical pathway. There is also evidence suggesting that such iron-sulfur clusters can function in coupling electron transfer to
10 proton transport. By binding Cys ligands from different subunits, iron-sulfur clusters effect dimer formation, as in the Fe protein of nitrogenase. Further, by straddling protein structural elements, iron-sulfur clusters are able to stabilize structures that are required for specific functions (eg. endonuclease III of *E. coli*). Proteins of ISPG's class have also been shown to protect proteins from the attack of intracellular proteases. Finally, proteins of ISPG's class are thought to be capable of
15 serving as storage devices for iron and possibly sulfide.

Thus, in a few examples, ISPG may be used advantageously as an iron or metal biosensor, for the treatment and/or diagnosis of iron overload disorders, or in applications involving stabilizing target proteins such as for protein production or for mediating protein interactions.

Additionally, structural aspects of ISPG suggest that it may be capable of mediating steroid
20 hormone synthesis, either in human or animals, or in engineered cell culture systems for the large scale production of hormones. ISPG may be used to act as an adrenal ferredoxin (known as adrenodoxin (ADX)), a vertebrate mitochondrial protein which transfers electrons from adrenodoxin reductase to cytochrome P450_{scc}, which is involved in cholesterol side chain cleavage. Its primary function as a soluble electron carrier between the NADPH-dependent adrenodoxin
25 reductase and several cytochromes P450 makes it an irreplaceable component of the steroid hormones biosynthesis in the adrenal mitochondria of vertebrates.

DRUG METABOLISM

Previous studies have revealed that cytochrome P-450 isozymes are responsible for drug metabolism, and oxidation by P-450 isozymes is a common aspect of the overall clearance of drugs.
30 Further studies have revealed that genetic polymorphism of cytochrome P-450 isozymes underlies a wide spectrum of substrates specificity in drug oxidation. In certain cases, genetic mutation and/or deletion of one critical isozyme gene results in a significant alteration of a phenotype projected on substrate specificity. It has been reported that CYP2D6 oxidizes more than 30 drugs (for example, M. Eichelbaum et al., *Pharmacol. Ther.*, Vol. 46, pp. 377-, 1990). Many anti-cancer drugs are
35 known to be oxygenated by cytochrome P450 enzymes to yield metabolites that are cytotoxic or cytostatic toward tumor cells. These include several commonly used cancer chemotherapeutic drugs, such as cyclophosphamide (CPA), its isomer ifosfamide (IFA), dacarbazine, procarbazine, thio-TEPA, etoposide, 2-aminoanthracene, 4-ipomeanol, and tamoxifen (LeBlanc, G. A. and

Waxman, D. J., *Drug Metab. Rev.* 20:395-439 (1989); Ng, S. F. and Waxman D. J., *Intl. J. Oncology* 2:731-738 (1993); Goeptar, A. R., et al., *Cancer Res.* 54:2411-2418 (1994); van Maanen, J. M., et al., *Cancer Res.* 47:4658-4662 (1987); Dehal, S. S., et al., *Cancer Res.* 57:3402-3406 (1997); Rainov, N. G., et al., *Human Gene Therapy* 9:1261-1273 (1998)). Bioreductive metabolism
5 that results in drug activation is also catalyzed by cytochrome P450 enzymes for a variety of anti-cancer drugs. Examples of such drugs include Adriamycin, mitomycin C, and tetramethylbenzoquinone (Goeptar, A. R., et al., *Crit. Rev. Toxicol.* 25:25-65 (1995); Goeptar, A. R., et al., *Mol. Pharmacol.* 44:1267-1277 (1993)). Those who have homozygous alteration in this recessive gene, are so-called "poor metabolizers (PMs)" and may suffer from severe side effects due
10 to poor metabolism of drugs (for example, see M. Eichelbaum et al., *Pharmacol. Ther.*, Vol. 46, pp. 377-, 1990). Such genetic alterations occur at rates of from 1 to 30% in different ethnic populations (for example, L. M. Distlerath et al., *J. Biol. Chem.*, Vol. 260, pp. 9057-, 1985).

The ISPG protein of the invention is thought to be capable of functioning as a soluble electron carrier in the electron transport chain involving one or more of the several available
15 cytochromes P450 enzymes. The ISPG protein may thus be useful in methods of killing neoplastic cells involving P450 (and ISPG) gene transfer and the use of bioreductive drugs that are activated by cytochrome P450, and in methods for evaluating the susceptibility of a sample compound to metabolism with respect to a specific cytochrome P-450 isozyme system.

Thus, in a first aspect, a drug activation/gene therapy strategy has been developed based on
20 a cytochrome P450 gene ("CYP" or "P450") in combination with a cancer chemotherapeutic agent that is activated through a P450-catalyzed monooxygenase reaction (Chen, L. and Waxman, D. J., *Cancer Research* 55:581-589 (1995); Wei, M. X., et al., *Hum. Gene Ther.* 5:969-978 (1994); U.S. Pat. No. 5,688,773, issued Nov. 18, 1997). Presently known drug-enzyme combinations can utilize established chemotherapeutic drugs widely used in cancer therapy. Such methods to obtain
25 enhanced chemosensitivity have been demonstrated both in vitro and in studies using a subcutaneous rodent solid tumor model and human breast tumor grown in nude mice in vivo, and is strikingly effective in spite of the presence of a substantial liver-associated capacity for drug activation in these animals (Chen, L., et al., *Cancer Res.* 55:581-589 (1995); Chen, L., et al., *Cancer Res.* 56:1331-1340 (1996)). The P450-based approach also shows significant utility for gene
30 therapy applications in the treatment of brain tumors (Wei, M. X., et al., *Human Gene Ther.* 5:969-978 (1994); Manome, Y., et al., *Gene Therapy* 3:513-520 (1996); Chase, M., et al., *Nature Biotechnol.* 16:444-448 (1998)).

Although the P450/drug activation system has shown great promise against several tumor types, further enhancement of the activity of this system is needed to achieve clinically effective,
35 durable responses in cancer patients. This requirement is necessitated by two characteristics that are inherent to the P450 enzyme system: (1) P450 enzymes metabolize drugs and other foreign chemicals, including cancer chemotherapeutic drugs, at low rates, with a typical P450 turnover number (moles of metabolite formed/mole P450 enzyme) of only 10-30 per minute; and (2) P450 enzymes metabolize many chemotherapeutic drugs with high K_m values, typically in the millimolar

range. This compares to plasma drug concentrations that are only in the micromolar range for many chemotherapeutic drugs, including drugs such as CPA and IFA. Thus, current approaches to P450 gene therapy may result in intratumoral drug activation at a low absolute rate and under conditions that are not saturating with respect to drug substrate. Furthermore, since P450 is expressed at a very high level in liver tissue, only a very small fraction of the administered chemotherapeutic drug is metabolized via the tumor cell P450 gene product using the currently available methods for P450 gene therapy (Chen, L. and Waxman, D. J., Cancer Res. 55:581-589 (1995)). As described in U.S. Patent No. 6,207,648, one enhancement involves introducing a P450 reductase (RED) gene in combination with a cytochrome P450 gene (and thus a P450 gene product) into neoplastic cells, the enzymatic conversion of a P450-activated chemotherapeutic drug to its therapeutically active metabolites is greatly enhanced within the cellular and anatomic locale of the tumor, thereby increasing both the selectivity and efficiency with which neoplastic cells are killed.

In a preferred embodiment, further enhancements to known prodrug-enzyme strategies may be achieved by introducing an ISPG gene into neoplastic cells, either alone or in combination with a P450 gene and/or a P450 reductase gene. Suitable vectors for the introduction and expression of said ISTG and P450 genes are known to one of skill in the art. The introduction of the ISPG gene and subsequent expression of the ISPG gene product may increase the enzymatic conversion of a P450-activated chemotherapeutic drug to its therapeutically active metabolites. Thus, the invention comprises a method for killing neoplastic cells comprising: (a) infecting the neoplastic cells with a vector for gene delivery, the vector comprising an ISPG gene capable of mediating enzymatic conversion of a chemotherapeutic agent by a P450 enzyme; (b) optionally infecting the neoplastic cells with a vector for gene delivery, the vector comprising a cytochrome P450 gene and/or a gene encoding RED; (b) treating the neoplastic cells with a chemotherapeutic agent that is activated by the product of the cytochrome P450 gene; and (c) killing the neoplastic cells.

The present invention also provides a reagent composition for use in evaluating drug metabolism by a specific cytochrome P-450 isozyme, which comprises a liver microsome lacking said specific P-450, said specific P-450 isozyme and a carrier material. The liver microsome may be of human source lacking CYP2D6, CYP2C 19, or CYP2A6. The CYP2D6 isozyme, CYP2C 19 isozyme, and CYP2A6 isozyme to be added may be a recombinant CYP2D6-expressing microsome, a recombinant CYP2C 19-expressing microsome, or a recombinant CYP2A6-expressing microsome. The reagent composition may comprise more than one kind of PM microsomes.

According to the present invention, there can be provided a reagent composition and a method for accurately quantitating the contribution of certain P-450 isozymes such as CYP2D6, CYP2C 19, and CYP2A6 in drug metabolism. The present invention provides a method for evaluating the susceptibility of a sample compound to metabolism with respect to a specific cytochrome P-450 isozyme, which comprises contacting the sample compound with a reagent composition prepared by adding said specific cytochrome P-450 isozyme and an ISPG protein to liver microsomes lacking said specific cytochrome P-450 isozyme in a carrier material. ISPG

would be useful in order to enhance efficiency of the P-450 isozyme in drug metabolism, thereby effectively amplifying the power of the assay to detect the contribution of a particular P-450 enzyme. In another embodiment the contribution to drug metabolism of a particular P-450 system can be assessed by focussing on a particular iron-sulfur protein associated with said specific P-450 system. In this aspect, the method comprises contacting the sample compound with a reagent composition prepared by adding said specific ISPG protein to liver microsomes lacking said ISPG protein in a carrier material. The method may further comprise (a) incubating a mixture of the sample compound and the reagent composition; (b) extraction of the reaction mixture obtained in Step (a); and (c) analyzing the reaction products isolated in Step (b). For the purposes of quantitating the assay, a plurality of the reagent compositions having different amount of the specific P-450 isozyme or ISPG protein may be subjected to Step (a) to (c), respectively. For example, the specific P-450 isozyme to be used in the method may be selected from CYP2D6, CYP2C19, CYP2A6, CYP1A1 and CYP2E1.

IRON BIOSENSOR

Iron-sulfur clusters have been found to serve as sensors of iron, dioxygen, superoxide ion and possibly nitric oxide. Two main mechanisms of sensing have been described. In one example, the oxidation $[\text{Fe}^2\text{S}_2]^{1+} \rightarrow [\text{Fe}^2\text{S}_2]^{2+}$ by dioxygen provides the signal for activation of a defense mechanism against superoxide, as observed with the SoxR protein of *E. coli*, and thus may serve a cytoprotective function (eg. useful for treatment of ischemia, etc.). In an alternative mechanism, oxidative disassembly or reassembly of a cluster provides the controlling signal as in the FNR protein of *E. coli*.

In a further detailed example, the ISPG protein of the invention may be used as an iron biosensor. An example of an iron biosensor is provided in U.S. Patent No. 5,516,697 (Kruzel et al.). In summary, ISPG can be immobilized in the vicinity of a device to measure the change in pH, causing a detectable variation of the potential upon the binding of iron.

In the process of sequestering iron, the sensing element, the ISPG protein is expected to release a number of protons of hydrogen (H^+) directly proportional to the atoms of iron bound. The release of protons during the binding of iron by ISPG becomes the operative feature which is measured by the biosensors. The release of protons causes a change in pH and is measured by an ion-selective field effect transistor or by pH sensitive paper. A sample containing iron is placed into a buffered solution, usually water. The sample may be diluted one or more times. In one embodiment of the sensors of the present invention, the release of protons is measured as the variation of the potential on the surface of an ion-selective field effect transistor (an ISFET) (Reviewed in: Biosensor Technology, edited by Buck et al. and published by Marcel Decker, Inc., 1990, entitled "Solid State Potentiometric Sensors" by Jiri Janata, pp 17-34). In another embodiment of the present invention, the protons released upon binding of iron by ISPG are detected by the change in pH using pH sensitive paper. Preferably, the iron selective element (ISPG) is incorporated in close proximity or integrated with the signal transducer, to give a

reagentless sensing system for iron. Since the signal can be amplified, only small quantities of ISPG are needed for detection of iron. The ISFET is modified by immobilizing ISPG on the surface of the ISFET or by a disposable membrane with immobilized ISPG that is in close proximity to the ISFET by attaching the ISPG -modified membrane to the surface of the ISFET. A sample containing iron, 5 for example a biological sample such as body fluid from a mammal, particularly a human, is then contacted with the ISPG -modified ISFET.

In order to produce an ISPG-modified ISFET as an independent sensor, an existing system which uses an ISFET designed to measure pH can be modified. Systems which presently use an ISFET to measure pH are the Sentron 2001 pH system, manufactured by Integrated Sensor 10 Technology, Federal Way, Wash.; the Corning 360i pH system, manufactured by Corning Incorporated, Corning N.Y.; or Orion 610 pH system, manufactured by Orion Analytical Technology, Inc., Boston, Mass. The modification required to measure the amount of iron in a sample is either to place an immobilized layer of ISPG on the ISFET of such a system or, alternatively, to provide a ISPG -modified membrane, i.e. a membrane coated with ISPG, which 15 will be in close proximity to the existing ISFET so as to detect the release of protons when the ISPG binds iron in a sample and records the change in potential.

Iron sensitive biosensors (as well as treatments for iron overload disorders) are extremely valuable. It is estimated that 30,000,000 Americans suffer from different types of iron related disorders, including a substantial proportion with profound iron deficiency syndrome. Detection of 20 bioaccessible iron is one of the most important measurements that doctors can use for early detection of iron deficiency, iron overload or other types of immunological disorders. To date iron is measured through a combination of blood tests that detect iron and iron binding capacity of transferrin, the protein that transports iron through the body. The current technology involves very sophisticated instrumentation which make this analysis prohibitively expensive and often requires 25 qualified personnel to analyze the sample. Therefore, there is a need for the direct assay of iron that combines simplicity and economics. ISPG may therefore be advantageously used in development of a biosensor for detecting the amount of iron in a sample.

STEROID BIOSYNTHESIS

As noted above, ISPG may be used as an adrenal ferredoxin (known as adrenodoxin 30 (ADX)), a vertebrate mitochondrial protein which transfers electrons from adrenodoxin reductase to cytochrome P450_{scc}, which is involved in cholesterol side chain cleavage and is an irreplaceable component of the steroid hormones biosynthesis in the adrenal mitochondria.

In therapeutic embodiments, ISPG may have particular importance in treatment of disorders where it is desired to increase the level of steroid hormone synthesis. As P450_{scc} has a critical role 35 in synthesis of the conversion of cholesterol into pregnenolone, ISPG may be used as a limiter or enhancer of steroid synthesis. In but one example, evidence has been shown that cytochrome P450_{scc} activity in the human placenta is limited by the supply of electrons to the P450_{scc}. Furthermore, Tuckey et al. Eur. J. Biochem. 1999 Jul;263(2): 319-325 have shown that p450_{scc}

activity can be increased considerably by adding adrenodoxin reductase and adrenodoxin. Thus, ISPG may be useful in the treatment of reproductive disorders by augmenting the electron supply to P450_{scc}, and thus increasing the level of progesterone synthesis. Accordingly, in another example, ISPG may be used to limit steroid synthesis, whether for therapeutic or for research uses.

- 5 ISPG may also be used in biological steroid synthesis processes for the production of steroid hormones. For example, Duport et al, Nat Biotechnol 1998 Feb;16(2):186-9 report a system for self-sufficient biosynthesis of pregnenolone and progesterone in engineered yeast wherein the first two steps of the steroidogenic pathway were reproduced in *Saccharomyces cerevisiae*. Engineering of sterol biosynthesis by disruption of the delta 22-desaturase gene and introduction of
- 10 the *Arabidopsis thaliana* delta 7-reductase activity and coexpression of bovine side chain cleavage cytochrome P450, adrenodoxin, and adrenodoxin reductase, lead to pregnenolone biosynthesis from simple carbon source. As ISPG is thought to be capable of functioning as an adrenodoxin protein, ISPG may be used as a function substitute in the system of Duport et al for adrenodoxin.

15 MTG (METALLOTHIONEIN) (Clone ID:654627)

- SEQ ID NOS 96 and 413 and clone FL11:654627_182-5-3-0-F10-F encode the polypeptide of SEQ ID NOS:265 and 527 respectively, a metallothionein protein which binds heavy metal. Said polypeptide of the invention is also referred herein as MTG. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOS:96 and 413 and polypeptides of SEQ
- 20 ID NO:265 and 527, described throughout the present application also pertain to the human cDNA of clone 654627, and the polypeptides encoded thereby.

- Metallothioneins (MT) [1,2,3] are small proteins which bind heavy metals such as zinc, copper, cadmium, nickel, etc., through clusters of thiolate bonds. MT's occur throughout the animal kingdom and are also found in higher plants, fungi and some prokaryotes and are thought to
- 25 play a role in metal detoxification or in the metabolism and homeostasis of metals. On the basis of structural relationships MT's have been subdivided into three classes. Class I includes mammalian MT's as well as MT's from crustacean and molluscs, but with clearly related primary structure. Class II groups together MT's from various species such as sea urchins, fungi, insects and cyanobacteria which display none or only very distant correspondence to class I MT's.
- 30 Class III MT's are atypical polypeptides containing gamma-glutamylcysteinyl units.

- Vertebrate class I MT's such as the MTG protein of the invention are proteins of typically 60 to 68 amino acid residues, 20 of these residues are cysteines that bind to 7 bivalent metal ions. As a signature pattern we chose a region that spans 19 residues and which contains seven of the metal-binding cysteines, this region is located in the N-terminal section of class-I MT's. A
- 35 consensus pattern for class I MT's is as follows: C-x-C-[GSTAP]-x(2)-C-x-C-x(2)-C-x-C-x(2)-C-x-K.

The MTG protein of SEQ ID NO 527 has a metallothionein domain (Prosite ref. PS00203) at amino acid positions 13-31; an N-glycosylation site at position 4 (NCSC), a protein kinase C

phosphorylation site at amino acid positions 18 (SCK), 28 (SCK) and 55 (SQR); a casein kinase II phosphorylation site at amino acid position 41 (TLVD); an N-myristoylation site at amino acid positions 10 (GVSTCTC); and a prokaryotic membrane lipoprotein lipid attachment site at amino acid position 3 (PNCSCAAGVSC).

5 THERAPEUTICS

Discovery of new proteins related to metallothioneins, and the polynucleotides that encode them, satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in diagnosing and treating heavy metal toxicity, cancer, inflammatory disease and immune disorders.

Acute or chronic exposure to heavy metals such as lead, arsenic, mercury or cadmium leads
10 to a variety of diseases and disorders involving neuromuscular, CNS, cardiovascular, and gastrointestinal effects. MTs may play a role in the prevention or alleviation of these conditions. In addition, MTs are transcriptionally regulated by glucocorticoids, which suggests that MTs have a direct role in the effects of glucocorticoids to treat inflammatory disease, immune disorders, and cancer. It is therefore thought that MTG may have important applications in the treatment of
15 inflammatory disease, immune disorders, and cancer as well as in cytoprotection in a variety of therapeutic applications.

In one preferred example, the MTG nucleic acids and protein may be used for suppressing the production of sunburn cells which is applicable in various manners with minimal adverse side effects, a method of inducing metallothionein, a method of treating skin diseases and a method of
20 screening ultraviolet rays, and further relates to cosmetic compositions and UV screening compositions.

Conventionally, steroids and zinc oxide formulations have been topically used as medicines for treating skin diseases such as dermatitis, sunburn, neurodermatitis, eczema and anogenital pruritus. Steroids, however, have been difficult to administer in large quantities for a prolonged
25 period due to their strong adverse side effects. Zinc oxide formulations, which have local astringent action, involve problems with respect to the manufacture of pharmaceuticals, since they are insoluble in water and are not usually administered internally.

Zinc, one of the indispensable trace metals in the living body, is known to participate in the development of sexual organs, promotion of wound healing and is also known to be a component of
30 a metalloenzyme, an accelerator for dehydrogenase, and to have various functions such as activating the immune system. Zinc is further known to be an inducing factor of metallothionein (MT), a metal-combining protein. It is reported that MT functions as a scavenger of free radicals which are generated at the onset of inflammations ["Dermatologica", Hanada, k., et al., 179 (suppl. 1) 143 (1989)].

35 As proposed in U.S. Patent No. 5,582,817 (Otsu et al), MTG may be useful in treatment of dermatological inflammations caused by external irritative stimulants, such as sunburn or the like, where MTG could act to quench the free radicals released from leukocytes, especially granulocytes which gather at the inflamed region, and thereby exhibit an anti-oxidation action to diminish cell

damage, especially to normal lymphocytes, to activate the immune system and further to prevent the accelerated aging of the skin. Formation of sunburn cells (SBCs) could be suppressed by administering zinc for inducing MTG to be present, or to increase MTG in the epidermal keratinous layer. Anti-oxidation action of MTG can also be useful in the treatment of skin problems resulting from radiation therapy by X rays, alpha rays, beta rays, gamma rays, neutron rays and accelerated electron rays.

Various zinc compounds have been studied by Otsu et al (supra) with respect to their pharmacological activities, who reported that zinc salts or zinc complexes of a certain compound have an excellent action of inducing metallothionein (MT) and suppressing sunburn cell (SBC) production due to UV rays, and thereby useful as components of cosmetic compositions or medicines for purposes of ameliorating sunburn, preventing sunburn, ameliorating sufferings from skin diseases and ameliorating other radiation induced disorders, leading to completion of the invention.

There are two different types of dermatological reactions caused by sunlight, one is an acute inflammatory change in the skin called sunburn, and the other is a subsequent melanin pigmentation called suntan. The light having a wave length in the range of 320 nm or less, called UVB, induces sunburn and is responsible for erythematous change. The erythemic reaction caused by UV rays, as opposed to a burn injury, does not occur immediately after the exposure to the sunlight, but rather occurs after a latent period of several hours. When sunburned skin is histopathologically examined, various degrees of inflammatory changes are recognized in the epidermis and dermis depending on the dose of radiation. Among such changes, a notable one is the generation of so-called sunburn cells (SBC) in the epidermis. A histologically stained tissue sample presents strongly and acidophilically stained cells which have pyknotic nuclei. This phenomenon indicates the necrosis of epidermal cells ("Fragrance Journal", 9, 15-20 (1991)). In order to prevent sunburn, para-aminobenzoic acid derivatives, cinnamic acid derivatives or the like UV absorbers mentioned above are used, but their UV absorbing effects are not necessarily satisfactory. What is more, they raise problems of cumbersome handling upon use, poor stability, low compatibility with other components of the composition, and also involve unsolved problems in water-resistance and oil-resistance.

In the field of medicines for the treatment of skin diseases, development of medicines which have minimal adverse side effects, and which have novel functions obtainable by both external and internal administrations has been desired. Also, in the field of the therapy and prevention of radiation disorders, medicines which can suppress and cure the disorders caused by oxidative reactions have been desired. Lastly, in the field of the manufacture of cosmetics, cosmetics which overcome the above-mentioned problems such as handling upon use and stability of the composition have been desired. Accordingly, the present invention encompasses providing therapeutic agents for treating skin diseases having the above-mentioned characteristics, wherein said agents are capable of inducing MTG for suppressing the formation of sunburn cells, and for use in cosmetic compositions. Also encompassed are methods of screening for therapeutic agents for

treating skin diseases comprising bringing a test compound into contact with a cell, tissue or animal model of disease, and detecting induction of MTG expression or function.

GENE EXPRESSION SYSTEMS

The MTG nucleic acid and proteins of the invention may also be advantageously used in the
5 production of recombinant proteins as biopharmaceutical products at commercial scale.

Previously, genes have been extensively expressed in mammalian cell lines, particularly in mutant Chinese Hamster Ovary (CHO) cells deficient in the dihydrofolate reductase gene (dhfr) as devised by the method of Urlaub et al, PNAS U.S.A. 77, 4216-4220, 1980. A variety of expression systems have been used. Many vectors for the expression of genes in such cells are therefore
10 available. Typically, the selection procedures used to isolate cells transformed with the expression vectors rely on using methotrexate to select for transformants in which both the dhfr and the target genes are coamplified. The dhfr gene, which enables cells to withstand methotrexate, is usually incorporated in the vector with the gene whose expression is desired. Selection of cells under increasing concentrations of methotrexate is then performed. This leads to amplification of the
15 number of dhfr genes present in each cell of the population, as cells with higher copy numbers withstand greater concentrations of methotrexate. As the dhfr gene is amplified, the copy number of the gene of interest increases concomitantly with the copy number of the dhfr gene, so that increased expression of the gene of interest is achieved. Unfortunately, these amplified genes have been reported to be variably unstable in the absence of continued selection (Schimke, J. Biol. Chem.
20 263, 5989-5992, 1988). This instability is inherent to the presently available expression systems of CHO dhfr.sup.- cells. For many years, several promoters have been used to drive the expression of the target genes such as the SV40 early promoter, the CMV early promoter and the SR.alpha. promoter. The CMV and SR.alpha. promoters are claimed to be the strongest (Wenger et al, Anal. Biochem. 221, 416-418, 1994).

25 In one report, the .beta.-interferon promoter has also been used to drive the expression of the .beta.-interferon gene in the mutant CHO dhfr.sup.- cells (U.S. Pat. No. 5,376,567). In this system, however, the selected CHO dhfr⁻ cells had to be superinduced by the method of Tan et al (Tan et al, PNAS U.S.A. 67, 464-471, 1970; Tan et al, U.S. Pat. No. 3,773,924) to effect a higher level of .beta.-interferon production. In this system a significant percentage of the superinduced
30 .beta.-interferon produced by the CHO dhfr⁻ cells was not glycosylated. The mouse metallothionein gene (mMT1) promoter has also been used for the expression of beta-interferon genes in CHO cells, BHK and LTK.sup.- mouse cells (Reiser et al 1987 Drug Res. 37, 4, 482-485). However, the expression of .beta.-interferon with this promoter was not as good as the SV40 early promoter in CHO cells. Further, .beta.-interferon expression from these cells mediated by the mMT1 promoter
35 was inducible by heavy metals. Heavy metals are however extremely toxic to the cells and this system was therefore abandoned. Instead, Reiser et al used the CHO dhfr⁻ expression system in conjunction with the SV40 early promoter (Reiser et al, Drug Res. 37,4, 482-485 (1987) and EP-A-

0529300) to produce .beta.-interferon in CHO dhfr- cells as derived by the method of Urlaub et al (1980).

As described in U.S. Patent No. 6,207,146 (Tan et al) beta.-interferon was expressed in wild-type CHO cells using a metallothionein based system. MTG may thus be used in similar applications so as to provide a system for expression of recombinant proteins. Tan et al demonstrates wild-type CHO cells transfected with a vector comprising a .beta.-interferon gene under the control of a mouse sarcoma viral enhancer and mouse metallothionein promoter (MSV-mMT1), a neo gene under the control of promoter capable of driving expression of the neo gene in both E. coli and mammalian cells and a human metallothionein gene having its own promoter. Transfected cells capable of expressing .beta.-interferon were selected by first exposing cells to geneticin (antibiotic G418) and thus eliminating cells lacking the neo gene and then exposing the surviving cells to increasing concentrations of a heavy metal ion.

The heavy metal ion enhanced the MSV-mMT1 promoter for the .beta.-interferon gene, thus increasing .beta.-interferon expression. The heavy metal ion also induced the human metallothionein gene promoter, causing expression of human metallothionein. The human metallothionein protected the cells against the toxic effect of the heavy metal ion. The presence of the heavy metal ion ensured that there was continual selection of cells which had the transfecting vector, or at least the .beta.-interferon gene and the human metallothionein gene and their respective promoters, integrated into their genome.

The selected cells that had been successfully transfected expressed .beta.-interferon. Expression was surprisingly improved when the cells were cultured in the presence of Zn^{2+} . The .beta.-interferon had improved properties, in particular a higher bioavailability, than prior .beta.-interferons.

These findings have general applicability and suggest that the MTG gene of the present invention may be used accordingly in expression systems. Accordingly, the present invention provides a nucleic acid vector comprising:

(i) a coding sequence which encodes a protein of interest and which is operably linked to a promoter capable of directing expression of the coding sequence in a mammalian cell in the presence of a heavy metal ion; (ii) a first selectable marker sequence which comprises an MTG gene of the invention and which is operably linked to a promoter capable of directing expression of the MTG gene in a mammalian cell in the presence of a heavy metal ion; and optionally (iii) a second selectable marker sequence which comprises a neo gene and which is operably linked to a promoter capable of directing expression of the neo gene in a mammalian cell;

CDPG (GLYCOSYL PHOSPHATIDYLINOSITOL-LINKED GLYCOPROTEIN) (Clone ID:1000902917)

SEQ ID NOS 3 and 341 and clone FL11:1000902917_223-52-4-0-G3-F encode the polypeptide of SEQ ID NOS 172 and 458 respectively, a glycosyl phosphatidylinositol-linked

glycoprotein protein which is thought to be a signal transducing polypeptide expressed in lymphoid, myeloid, and erythroid cells. Said polypeptide comprises a CD24 signal transducing domain as well as a GPI-anchor of the invention is also referred herein as CDPG. CDPG is believed to be highly glycosylated, and it is expected that CDPG molecular weight will vary among cell types and cell developmental stage due to differences in glycosylation patterns, providing further specificity in its use as a therapeutic target. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs: 3 and 341 and polypeptides of SEQ ID NO: 172 and 458, described throughout the present application also pertain to the human cDNA of clone 1000902917, and the polypeptides encoded thereby.

It is suggested that CDPG may have a specific role to play in early thymocyte development. The CDPG protein is thought to be extensively o-glycosylated may be capable of modulating b-cell activation responses. As a signaling transducer, the CDPG polypeptide's signal transducing function may in some embodiments be triggered by the binding of a lectin-like ligand to the CD24 domain carbohydrates, and the release of second messengers allowing signaling. The CDPG polypeptide is thought to have important functions in regulating the differentiation and/or growth of lymphoid, myeloid, and erythroid cells, including specifically promoting antigen dependent proliferation of b-cells, and preventing their terminal differentiation into antibody-forming cells.

Additionally, based on a growing body of evidence characterizing the CD24 domain function, it is proposed that CDPG may have a role as a potent stimulator of neurite outgrowth, and thus may be useful in the treatment of central nervous system disorders.

Fragments of CDPG may also be useful, eg. GPI-anchor domain for example in the development of soluble T-cell receptors (U.S. Patent No. 6,080,840) or any suitable application where a temporally controlled solubilization of a protein of interest is desired. For example, a fragment comprising the CDPG GPI-anchor domain can be used in the production of soluble molecules by replacing the transmembrane domains of the cDNA of a protein of interest with a sequence comprising the CDPG glycosylphosphatidyl inositol (GPI) linkage. These chimeric cDNAs are then transferred into an expression vector containing a strong promoter and a mutant (e.g. DHFR) gene allowing high levels of transcriptional expression and amplification of the gene. These chimeric genes are then cotransfected into a selected cell type, preferably lacking the endogenous protein of interest, and transfectants are selected and can also be screened with antibodies for the protein of interest. These GPI linked proteins of interest can then be solubilized by cleavage with the enzyme phosphatidyl inositol specific phospholipase C (PI-PLC) and purified/concentrated from the supernatant (e.g. by passage over a protein of interest-reactive antibody affinity column).

35 *THERAPEUTICS*

CDPG or inhibitors of CDPG may be used in the treatment of any disorder where it is desired to regulate B-cell proliferation or differentiation. CDPG or inhibitors thereof may be useful in the treatment of B-cell neoplasms, a heterogeneous group of diseases characterized by different

maturation states of the B-cell, which are related to the aggressiveness of the disorder. Chronic lymphocytic leukemia (CLL) is characterized by proliferation and accumulation of B-lymphocytic leukemia (BLL) is characterized by proliferation and accumulation of B-lymphocytes that appear morphologically mature but are biologically immature. This disorder accounts for 30% of leukemias in Western countries. The disorder is characterized by proliferation of biologically immature lymphocytes, unable to produce immunoglobulins, which cause lymph node enlargement. As a regulator of B-cell proliferation and differentiation, CDPG and/or inhibitors of CDPG may be useful for inhibiting proliferation of leukemic B-cells in CLL patients.

CDPG may also be useful in the modulation of cell growth in the CNS. CD24 is known to be highly expressed in neurons and has been demonstrated as capable of inhibiting neurite outgrowth of dorsal root ganglion neurons while promoting neurite outgrowth of cerebellar neurons via interaction with an L1 protein.

SELECTABLE CELL MARKERS

In one aspect, the CDPG polypeptide may be used as a selectable cell marker and to a method of using the selectable marker to identify a cell. Viruses such as recombinant retroviruses have been used as a vehicle for gene transfer based on their potential for highly efficient infection and non-toxic integration of their genome into a wide range of cell types. The transfer of exogenous genes into mammalian cells may be used, for example in gene therapy to correct an inherited or acquired disorder through the synthesis of missing or defective gene products in vivo. The expression of exogenous genes in cells may be useful in somatic gene therapy, to correct hereditary disorders at the level of the gene. Hemopoietic stem cells are particularly suited to somatic gene therapy as regenerative bone marrow cells may be readily isolated, modified by gene transfer and transplanted into an immunocompromised host to reconstitute the host's hemopoietic system.

Gene therapy involving bone marrow transplant with recombinant primary hemopoietic stem cells requires efficient gene transfer into the stem cells. As a very small number of primary stem cells can reconstitute the entire host hemopoietic system it is important that the transferred gene be efficiently expressed in the recombinant stem cells transferred. The transfer of foreign genes into a reconstituted host hemopoietic system has been limited by the availability of a selectable marker which permits the rapid and non-toxic selection of cells which are efficiently expressing the transferred gene. Currently available selection markers may not be suitable for primary hemopoietic stem cells since they may alter the proliferative ability or biological characteristics of the cells. The transfer of foreign genes into a reconstituted host hemopoietic system has also been limited by the availability of a viral vector capable of expression in hemopoietic stem cells, especially where more than one transcriptional unit is present in the vector (Botrell, D. R. L. et al., 1987, Mol. Biol. Med. 4:229).

U.S. Patent No. 5,804,177 (Humphries et al) has demonstrated that the cell surface protein CD24 (also M1/69-J11d heat stable antigen) can be used as a dominant marker in a recombinant viral vector. A nucleotide sequence encoding the cell surface protein CD24 in a recombinant viral

vector was used to infect hematopoietic stem cells and cells infected with the recombinant viral vector were rapidly and non-toxically selected for in vitro using fluorescence activated cell sorting (FACS), demonstrating a good correlation between proviral copy number and expression of selectable marker.

- 5 CD24 is a signal transducing molecule found on the surface of most human B cells that can modulate their responses to activation signals, and is structurally closely related to CDPG. The CD24 CDNA (approximately 300 bps) has been cloned (Kay, R. et al, 1991, J. Immunol. 147:1412) and encodes a mature peptide of only 31 to 35 amino acids that is extensively glycosylated and attached to the outer surface of the plasma membrane by a glycosyl phosphatidylinositol lipid anchor. M1/69-J11d heat stable antigen is a genetically similar homologous murine peptide widely
10 expressed on a variety of hemopoietic cell types (Kay, R. et al., 1990, J. Immunol. 145:1952).

- It is thus proposed that a recombinant viral vector can be used to successfully transfer and express the CDPG gene in primitive hemopoietic stem cells such that they are able to repopulate lethally irradiated recipients. Preferably foreign CDPG antigen expression in repopulated animals
15 persists post transplantation such that the biological function of the repopulated hemopoietic cells is not affected by the expression of the CDPG antigen. CDPG may subsequently be found to be expressed in any or all of hemopoietic lineages including granulocytes, macrophages, pro-erythrocytes, erythrocytes and T and B lymphocytes. Therefore, the cell surface protein CDPG may be particularly useful as a marker for hematopoietic stem cells capable of repopulation in vivo and
20 as a selectable marker in gene therapy. The recombinant viral vectors also have the advantage that the nucleotide sequence encoding the marker is very small, leaving a large amount of space for the insertion of additional genes of interest such as those coding for exogenous genes.

- In a preferred embodiment of the invention a recombinant viral vector is used to introduce the nucleotide sequence into the cell. Preferably, the CDPG nucleotide sequence is operatively
25 linked to one or more regulatory elements. The recombinant viral vector of the invention may be used as a marker for an exogenous gene to be expressed in a host cell. The invention further provides a method of identifying a cell and progeny thereof comprising: providing a cell; infecting the cell with a recombinant viral vector of the invention under suitable conditions to allow expression of the cell surface protein CDPG on the cell; and, identifying the cell and progeny
30 thereof by detecting expression of the cell surface protein CDPG on the cell or progeny thereof. Cells infected with a recombinant viral vector of the invention and expressing the cell surface protein may be transplanted into a host, and the cell and progeny thereof may be identified after transplantation by removing biological samples from the host, and assaying for cells expressing the cell surface protein. A recombinant viral vector of the invention may be directly introduced into a
35 host.

ENRICHING STEM CELL COMPOSITIONS

As it is proposed that CDPG is involved in early thymocyte development and is found on the cell surface due to its GPI-anchor, CDPG nucleic acids and polypeptides of the present

invention may also be used to obtain novel antibody compositions useful for preparing cell preparations containing human hematopoietic cells.

There is a continued interest in developing stem cell purification techniques. Pure populations of stem cells will facilitate studies of hematopoiesis. Transplantation of hematopoietic cells from peripheral blood and/or bone marrow is also increasingly used in combination with high-dose chemo- and/or radiotherapy for the treatment of a variety of disorders including malignant, nonmalignant and genetic disorders. Very few cells in such transplants are capable of long-term hematopoietic reconstitution, and thus there is a strong stimulus to develop techniques for purification of hematopoietic stem cells. Furthermore, serious complications and indeed the success of a transplant procedure is to a large degree dependent on the effectiveness of the procedures that are used for the removal of cells in the transplant that pose a risk to the transplant recipient. Such cells include T lymphocytes that are responsible for graft versus host disease (GVHD) in allogenic grafts, and tumour cells in autologous transplants that may cause recurrence of the malignant growth.

Hematopoietic cells have been separated on the basis of physical characteristics such as density and on the basis of susceptibility to certain pharmacological agents which kill cycling cells. The advent of monoclonal antibodies against cell surface antigens has greatly expanded the potential to distinguish and separate distinct cell types. There are two basic approaches to separating cell populations from bone marrow and peripheral blood using monoclonal antibodies. They differ in whether it is the desired or undesired cells which are distinguished/labeled with the antibody(s). In positive selection techniques the desired cells are labeled with antibodies and removed from the remaining unlabeled/unwanted cells. In negative selection, the unwanted cells are labeled and removed. Antibody/complement treatment and the use of immunotoxins are negative selection techniques, but FACS sorting and most batch wise immunoadsorption techniques can be adapted to both positive and negative selection. In immunoadsorption techniques cells are selected with monoclonal antibodies and preferentially bound to a surface which can be removed from the remainder of the cells e.g. column of beads, flasks, magnetic particles. Immunoadsorption techniques have won favor clinically and in research because they maintain the high specificity of targeting cells with monoclonal antibodies, but unlike FACS sorting, they can be scaled up to deal directly with the large numbers of cells in a clinical harvest and they avoid the dangers of using cytotoxic reagents such as immunotoxins, and complement.

Current positive selection techniques for the purification of hematopoietic stem cells target and isolate cells which express CD34. However, positive selection procedures suffer from many disadvantages including the presence of materials such as antibodies and/or magnetic beads on the CD34⁺ cells, and damage to the cells resulting from the removal of these materials.

Negative selection has been used to remove minor populations of cells from clinical grafts. These cells are either T-cells or tumour cells that pose a risk to the transplant recipient. The efficiency of these purges varies with the technique and depends on the type and number of

antibodies used. Typically, the end product is very similar to the start suspension, missing only the tumor cells or T-cells.

As described in U.S. Patent No. 5,877,299, Thomas et al developed a negative selection technique that uses an antibody composition containing antibodies specific for glycophorin A, CD3, CD24, CD16, CD14 and optionally CD45RA, CD36, CD2, CD19, CD56, CD66a, and CD66b, which reportedly gave a cell preparation highly enriched for human hematopoietic and progenitor cells. Maximum enrichment of early progenitor and stem cells (CD34⁺, CD38⁻ cells) was observed when anti-CD45R and anti-CD36 were included in the antibody composition. However, as CDPG is proposed as acting in early thymocyte development, CDPG may be used advantageously to develop more effective antibody compositions for selecting hematopoietic stem cells. Accordingly, the invention encompasses antibodies specific for CDPG polypeptides of the invention and antibody compositions comprising, consisting of or consisting essentially of antibodies specific for CDPG, glycophorin A, CD3, CD24, CD16, CD14 and optionally CD45RA, CD36, CD2, CD19, CD56, CD66a, and CD66b.

Use of the antibody composition comprising CDPG in a negative selection technique to prepare a cell preparation which is enriched for hematopoietic stem cells and progenitor cells may offer significant advantages over conventional techniques. The antibody composition is applied in one step to a sample of peripheral blood, bone marrow, cord blood or frozen bone marrow, preferably without additional enrichments steps which could result in loss of, or damage to, progenitor and stem cells.

PRSG (PROLINE-RICH CALCIUM-BINDING PROTEIN) and HSTG (BASIC HISTIDINE RICH SALIVARY GLAND PEPTIDE) (Clone ID:338112 and 1000839315 respectively)

SEQ ID NOS 22 and 358 and clone FL11:1000839315_220-26-1-0-F3-F encode the polypeptide of SEQ ID NOS 191 and 475 respectively, a basic histidine rich salivary gland peptide referred to herein as HSTG and expected to have potent antimicrobial properties. Preferably, the amino acid sequence of HSTG (SEQ ID NO 475) comprises a tyrosine at amino acid position 40. The HSTG protein also comprises a Pattern-DE: Protein kinase C phosphorylation site at amino acid position 51 (SSK). It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOS: 22 and 358 and polypeptides of SEQ ID NO: 191 and 475, described throughout the present application also pertain to the human cDNA of clone 1000839315, and the polypeptides encoded thereby.

SEQ ID NOS 8 and 345 and clone FL11:338112_174-1-1-0-A11-F encode the polypeptide of SEQ ID NOS 177 and 462 respectively, a proline-rich protein referred to herein as PRSG believed to be a component of saliva and a calcium binding protein also possessing potent antimicrobial properties. Preferably, the amino acid sequence of PRSG (SEQ ID NO 462) comprises a proline residue at amino acid position 96, an arginine residue at amino acid position 100, a glutamine residue at amino acid position 102, and/or a glycine residue at amino acid position

103. The PRSG protein also comprises a casein kinase II phosphorylation site at amino acid positions 15 (SAQD), 24 (SQED), and 59 (SAGD) and an N-myristoylation site at amino acid position 52 (GGQQSQ). It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:8 and 345 and polypeptides of SEQ ID NO: 177 and 462,
5 described throughout the present application also pertain to the human cDNA of clone 338112, and the polypeptides encoded thereby.

A study of saliva and its tooth-protective components reveals at least four important functions of saliva: (1) buffering ability, (2) a cleansing effect, (3) antibacterial action, and (4) maintenance of a saliva supersaturated in calcium phosphate. Several salivary constituents serve one
10 or more of these functions. Research has yielded important information about organic and inorganic secretory products. It is also clear that saliva as a unique biologic fluid has to be considered in its entirety to account fully for its effects on teeth. Saliva is greater than the sum of its parts. One reason for this is that salivary components display redundancy of function, each often having more than one function. This redundancy, however, does not imply that proteins that share functional
15 roles all contribute to the same degree. For instance, when comparing proteins that inhibit calcium phosphate precipitation, statherin and acidic proline-rich proteins are most potent, whereas histatins, cystatins, and mucins appear to play lesser roles. The complex interaction between proteins is another major factor contributing to saliva's function. In this regard, heterotypic complexes of various proteins have been shown to form on hydroxyapatite. Mucin binding to other salivary
20 proteins, including proline-rich proteins, histatins, cystatins, and statherin, is well documented. The complexes, whether adsorbed to the tooth surface or in saliva, have important implications for bacterial clearance, selective bacterial aggregation on the tooth surface, and control of mineralization and demineralization. Finally, proteolytic activity of saliva generates numerous products whose biologic activities are often different from their parent compounds. The ability of
25 saliva to deliver fluoride to the tooth surface constantly makes salivary fluoride an important player in caries protection largely by promoting remineralization and reducing demineralization. Saliva is well adapted to protection against dental caries. Saliva's buffering capability; the ability of the saliva to wash the tooth surface, to clear bacteria, and to control demineralization and mineralization; saliva's antibacterial activities; and perhaps other mechanisms all contribute to its essential role in
30 the health of teeth. The fact that the protective function of saliva can be overwhelmed by bacterial action indicates the importance of prevention and therapy as in other infectious diseases. With knowledge of salivary components and their interactions, the use of modified oral molecules as therapeutic agents may become an important contributor to oral health.

Proline-rich proteins are major components of parotid and submandibular saliva in humans
35 as well as other animals. They can be divided into acidic, basic and glycosylated proteins. The proline-rich proteins are apparently synthesized the acinar cells of the salivary glands and their phenotypic expression is under complex genetic control. The acidic proline-rich proteins will bind calcium with a strength which indicates that they may be important in maintaining the concentration of ionic calcium in saliva. Moreover they can inhibit formation of hydroxyapatite, whereby growth

of hydroxyapatite crystals on the tooth surface in vivo may be avoided. Both of these activities as well as the binding site for hydroxyapatite are located in the N-terminal proline-poor part of the protein.

Basic histidine rich salivary gland peptides such as the peptide of SEQ ID NO. 191 and 475, also referred to as histatins are a group of electrophoretically distinct histidine-rich polypeptides with microbicidal activity found in human parotid and submandibular gland secretions. Histatins 1, 3, and 5 are homologous proteins that consist of 38, 32, and 24 amino acid residues, respectively, that have been shown to kill the pathogenic yeast, *Candida albicans*. More recently histatins 2, 4, 6, and 7-12 were isolated and characterized Troxler RF et al, J Dent Res 1990 Jan;69(1):2-6. Histatin 2 was found to be identical to the carboxyl terminal 26 residues of histatin 1; histatin 4 was found to be identical to the carboxyl terminal 20 residues of histatin 3; and histatin 6 was found to be identical to histatin 5, but contained an additional carboxyl terminal arginine residue. The amino acid sequences of histatins 7-12 formally corresponded to residues 12-24, 13-24, 12-25, 13-25, 5-11, and 5-12, respectively, of histatin 3, but could also arise proteolytically from histatin 5 or 6. Troxler et al provides further guidance on the structural elements and relationship of histatins to one another in the context of their genetic origin, biosynthesis and secretion into the oral cavity, and potential as reagents in anti-candidal studies. The HSTG polypeptide and fragments thereof are therefore expected to have valuable properties and uses in antimicrobial applications, particularly in antifungal applications. Supporting such uses is a considerable body of evidence, including MacKay BJ et al, Infect Immun. 1984 Jun;44(3):695-701, Growth-inhibitory and bactericidal effects of human parotid salivary histidine-rich polypeptides on *Streptococcus* mutants; MacKay BJ et al, Infect Immun. 1984 Jun;44(3):688-94, Isolation of milligram quantities of a group of histidine-rich polypeptides from human parotid saliva; Pollock JJ et al, Infect Immun. 1984 Jun;44(3):702-7, Fungistatic and fungicidal activity of human parotid salivary histidine-rich polypeptides on *Candida albicans*; and Xu T et al, Infect Immun. 1991 Aug;59(8):2549-54, Anticandidal activity of major human salivary histatins. Furthermore, tissue distribution of RNAs for cystatins, histatins, statherin, and proline-rich salivary proteins in humans and macaques is further discussed in Sabatini et al, J Dent Res 1989 Jul;68(7):1138-45.

In further embodiments, the skilled artisan will appreciate that fragments and analogues of PRSG and HSTG may readily be generated and selected. Selection of preferred fragments and analogies may be carried out by assaying for a desired antimicrobial activity. For example, synthetic histatin analogues and methods for obtaining such analogies with broad-spectrum antimicrobial activity are described in Helmerhorst EJ et al, Biochem J 1997 Aug 15;326 (Pt 1):39-45, where histatin analogies inhibited the growth of the second most common yeast found in clinical isolates, *Torulopsis glabrata*, of oral- and non-oral pathogens such as *Prevotella intermedia* and *Streptococcus* mutants, and of a methicillin-resistant *Staphylococcus aureus*.

Thus, in preferred embodiments, the PRSG and/or HSTG polypeptides or fragments thereof may be used in oral, injectable, topical or edible compositions for the treatment of infection PRSG

and/or HSTG polypeptides may also be used as antimicrobial/antifungal compositions for disinfection of surfaces (e.g. in industrial settings).

In a preferred example further discussed below, the PRSG and/or HSTG polypeptides or fragments thereof are used in oral, topical (e.g. mouthwash) or edible compositions optionally containing additional salivary proteins to provide an anticaries effect. While there is an interest in developing and marketing products which reduce caries without reliance on a high level of fluoride ions (such as in fluoridated water and fluoride toothpastes), there have not been many reports of such approaches meeting with success. While certain cysteine-rich proteins have been proposed useful in the treatment of dental caries (U.S. Patent No. 5,688,766, Revis et al), the present invention provides PRSG and HSTG polypeptides which may provide higher potency, efficacy and range of disinfection and protection.

PRSG and HSTG polypeptide compositions can be administered in a formulation comprising a carrier. A preferred carrier composition for the active(s) of this invention are oral compositions. Such compositions include toothpastes, mouthrinses, liquid dentifrices, lozenges, chewing gums or other vehicle suitable for use in the oral cavity. Toothpastes and mouthrinses are the preferred systems. The abrasive polishing material contemplated for use in the toothpaste compositions of the present invention can be any material which does not excessively abrade dentin. These include, for example, silicas including gels and precipitates, calcium carbonate, dicalcium orthophosphate dihydrate, calcium pyrophosphate, tricalcium phosphate, calcium polymetaphosphate, insoluble sodium polymetaphosphate, hydrated alumina, and resinous abrasive materials such as particulate condensation products of urea and formaldehyde, and others such as disclosed by Cooley et al. in U.S. Pat. No. 3,070,510, Dec. 25, 1962, incorporated herein by reference. Mixtures of abrasives may also be used. Silica dental abrasives, of various types, can provide the unique benefits of exceptional dental cleaning and polishing performance without unduly abrading tooth enamel or dentin. For these reasons, they are preferred for use herein. Flavoring agents can also be added to toothpaste compositions. Suitable flavoring agents include oil of wintergreen, oil of peppermint, oil of spearmint, oil of sassafras, and oil of clove. Sweetening agents which can be used include aspartame, acesulfame, saccharin, dextrose, levulose and sodium cyclamate. Flavoring and sweetening agents are generally used in toothpastes at levels of from about 0.005% to about 2% by weight. Toothpaste compositions can also contain emulsifying agents. Suitable emulsifying agents are those which are reasonably stable and foam throughout a wide pH range, including non-soap anionic, nonionic, cationic, zwitterionic and amphoteric organic synthetic surfactants. Water is also present in the toothpastes of this invention. Water employed in the preparation of commercially suitable toothpastes should preferably be deionized and free of organic impurities. In preparing toothpastes, it is necessary to add some thickening material to provide a desirable consistency. Preferred thickening agents are carboxyvinyl polymers, carrageenan, hydroxyethyl cellulose and water soluble salts of cellulose ethers such as sodium carboxymethyl cellulose and sodium carboxymethyl hydroxyethyl cellulose. Natural gums such as gum karaya, xanthan gum, gum arabic, and gum tragacanth can also be used. Colloidal magnesium aluminum

silicate or finely divided silica can be used as part of the thickening agent to further improve texture. Thickening agents in an amount from 0.2% to 5.0% by weight of the total composition can be used. It is also desirable to include some humectant material in a toothpaste to keep it from hardening. Suitable humectants include glycerin, sorbitol, and other edible polyhydric alcohols at a level of from about 15% to about 70%.

Another preferred embodiment of the present invention is a mouthwash composition. Conventional mouthwash composition components can comprise the carrier for the agents of the present invention. Mouthwashes generally comprise from about 20:1 to about 2:1 of a water/ethyl alcohol solution or be alcohol free and preferably other ingredients such as flavor, sweeteners, humectants and sudsing agents such as those mentioned above for dentifrices. The humectants, such as glycerin and sorbitol give a moist feel to the mouth. Generally, on a weight basis the mouthwashes of the invention comprise 0% to 60% (preferably 5% to 20%) ethyl alcohol, 0% to 20% (preferably 5% to 20%) of a humectant, 0% to 2% (preferably 0.01% to 1.0%) emulsifying agents, 0% to 0.5% (preferably 0.005% to 0.06%) sweetening agent such as saccharin or natural sweeteners such as stevioside 0% to 0.3% (preferably 0.03% to 0.3%) flavoring agent, and the balance water.

The pH of the present compositions and/or the pH in the mouth can be any pH which is safe for the mouth's hard and soft tissues. Such pH's are generally from about 3 to about 10, preferably from about 4 to about 8. Other acceptable oral carriers include gums, lozenges, as well as other forms. Such suitable forms are disclosed in U.S. Pat. No. 4,083,955, Apr. 11, 1978 to Grabenstetter et al. incorporated herein in its entirety by reference. Edible compositions are also suitable for use as the carrier compositions herein. Edible compositions include many types of solid as well as liquid compositions. Such compositions include, for example, soft drinks, citrus drinks, cookies, cakes, breads among many others. Such compositions may contain sugar or another sweetener, water, flour, shortening, other fibers such as wheat, corn, barley, rye, oats, psyllium and mixtures thereof.

HSDG (HYDROXYSTEROID DEHYDROGENASE) (Clone ID:495917)

SEQ ID NOS 54 and 381 and clone FL11:495917_160-22-4-0-D8-F encode the polypeptide of SEQ ID NOS 223 and 498 respectively, a hydroxysteroid dehydrogenase referred to herein as HSDG. As the HSDG polypeptide is implicated in steroid hormone regulation, preferably glucocorticoid metabolism, HSDG may be useful in any applications where steroid hormones levels are to be increased or inhibited. HSDG may be useful in the treatment of disease treatable by steroid hormones. HSDG inhibitors may also be useful in systems for self-sufficient biosynthesis of steroid hormones. HSDG inhibitors of endogenous HSDG activity may allow the recovery of higher amounts of glucocorticoids and/or other synthesized steroid hormones from these cell systems.

It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs: 54 and 381 and polypeptides of SEQ ID NO: 223 and 498, described throughout the present application also pertain to the human cDNA of clone 495917, and the polypeptides encoded thereby.

- 5 The HSDG polypeptides of the invention comprise leucine zipper pattern (Prosite ref. PS00029) at amino acid positions 58-79 as well as a N-myristoylation site at amino acid position 36 (GANAGV) of SEQ ID NO 498. In one example, HSDG polypeptides may be used in the production or therapeutic modulation of glucocorticoids. The skilled artisan will recognize that any suitable HSDG polypeptides or variants or fragments thereof capable of metabolizing
- 10 glucocorticoids can readily be used.

- In one embodiment HSDG activity can be determined by detecting levels of glucocorticoids or metabolites or glucocorticoids. Moreover, structural aspects of 11 β -hydroxysteroid dehydrogenase have been documented in the art and may serve as a guidelines for developing suitable HSDG variants and fragments. The HSDG polypeptides of the invention may allow
- 15 modulation of glucocorticoid activity or identification (e.g. drug screening) of compounds capable of specifically modulating glucocorticoid levels or activity. HSDG may be useful in allowing the modulation of steroid hormone synthesis, or glucocorticoid synthesis to be carried in a tissue specific manner, thereby offering improved methods for treating disease with decreased risk of side effects,

- 20 Corticosteroids, also referred to as glucocorticoids are steroid hormones, the most common form of which is cortisol. Modulation of glucocorticoid activity is important in regulating physiological processes in a wide range of tissues and organs. Glucocorticoids act within the gonads to directly suppress testosterone production (Monder, C et al, (1994) Steroids 59, 69-73). High levels of glucocorticoids may also result in excessive salt and water retention by the kidneys,
- 25 producing high blood pressure.

- Glucocorticoid action is mediated via binding of the molecule to a receptor, defined hereinafter as either a mineralocorticoid receptor (MR) or a glucocorticoid receptor (GR). Krozowski, Z. S. et al, ((1983) Proc. Natl. Acad. Sci. USA 80, 6056-6060) and Beaumont, K. et al, ((1983) Endocrinology 113, 2043-2049) showed that MR of adrenalectomised rats have an equal
- 30 affinity for the mineralocorticoid aldosterone and glucocorticoids, for example corticosterone and cortisol. Confirmatory evidence has been found for human MR (Arriza, J. L et al, (1988) Neuron 1, 887-900). In patients suffering from the congenital syndrome of Apparent Mineralocorticoid Excess (AME), cortisol levels are reportedly elevated and bind to and activate MRs normally occupied by aldosterone, the steroid that regulates salt and water balance in the body. Salt and water
- 35 are retained in AME patients causing severe hypertension.

Like HSDG, the enzyme 11 β -hydroxysteroid dehydrogenase (11 β HSD), also discussed in U.S. Patent No. 5,965,372 (Funder et al) may be involved in converting glucocorticoids into metabolites that are unable to bind to MRs (Edwards et al, (1988) Lancet. 2: 986-9; Funder et al, (1988) Science 242, 583,585), present in mineralocorticoid target tissues, for example kidney,

pancreas, small intestine, colon, as well as the hippocampus, placenta and gonads. For example, in aldosterone target tissues 11β HSD inactivates glucocorticoid molecules, allowing the much lower circulating levels of aldosterone to maintain renal homeostasis. When the 11β HSD enzyme is inactivated, for example in AME patients or following administration of glycyrrhetic acid, a component of licorice, severe hypertension results. Further, placental 11β HSD activity may protect the foetus from high circulating levels of glucocorticoid which may predispose to hypertension in later life (Edwards et al., 1993). Biochemical characterisation of activity has indicated the presence of at least two 11β HSD isoenzymes (11β HSD1 and 11β HSD2) with different cofactor requirements and substrate affinities. The 11β HSD1 enzyme is a low affinity enzyme that prefers NADP⁺ as a cofactor (Agarwal et al., 1989). The 11β HSD2 enzyme is a high affinity enzyme (K_m for glucocorticoid=10 nM), requiring NAD⁺, not NADP⁺ as the preferred cofactor, belonging to a class of glucocorticoid dehydrogenase enzymes hereinafter referred to as "NAD⁺ dependent glucocorticoid dehydrogenase" enzymes.

Inverse correlation between 11β HSD enzyme activity in human granulosa-lutein cells and the success of IVF has further been shown, suggesting that activity of this enzyme might be related to the success of embryo attachment and implantation following IVF. The measurement of ovarian 11β HSD enzyme activity as a prognostic indicator for the outcome of assisted conception in all species, is the subject of UK Patent Application No 9305984. However, the disclosure of Michael et al. ((1993) Lancet 342, 711-712), and corresponding UK Patent Application No 9305984 do not identify, or even suggest which isoenzyme in the ovary might be a predictive indicator of IVF embryo transfer, or a means of distinguishing isoenzymes of 11β HSD in the prediction of IVF embryo transfer outcomes. In fact, the enzyme assay procedure might detect all isoenzymes of 11β HSD activity in the cell, some of which may be hitherto uncharacterised.

Thus, the human HSDG hydroxysteroid dehydrogenase enzyme and the nucleic acids encoding it providing novel means for the development of gene therapies and identification of HSDG activators and inhibitors which alter the endogenous activity of this hydroxysteroid dehydrogenase enzyme in a cell. The present invention also permits the screening, through genetic or immunological means, levels of expression of genes encoding the NAD⁺ dependent glucocorticoid dehydrogenase enzyme in various tissue or organ types, including for example, skin, colon, kidney, placenta, and gonads, amongst others.

S100G CALCIUM BINDING PROTEIN (Clone ID:200895)

The polynucleotides of clone FL11:200895_116-055-1-0-H11-F and SEQ ID NOS 132 and 428 encode for the S100 calcium binding protein referred to herein as S100G. SEQ ID NOS 301 and 542 provide the amino acid sequence corresponding to the nucleic acid sequences of SEQ ID NOS 132 and 428, respectively. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOS:132 and 428 and polypeptides of SEQ ID NO: 301 and 542,

described throughout the present application also pertain to the human cDNA of clone200895, and the polypeptides encoded thereby.

BACKGROUND

In nearly all eukaryotic cells, calcium (Ca^{2+}) functions as an intracellular signaling molecule in diverse cellular processes including cell proliferation and differentiation, neurotransmitter secretion, glycogen metabolism, and skeletal muscle contraction. Within a resting cell, the concentration of Ca^{2+} in the cytosol is extremely low, $<10^{-7}$ M. However, when the cell is stimulated by an external signal, such as a neural impulse or a growth factor, the cytosolic concentration of Ca^{2+} increases by about 50-fold. This influx of Ca^{2+} is caused by the opening of plasma membrane Ca^{2+} channels and the release of Ca^{2+} from intracellular stores such as the endoplasmic reticulum. Ca^{2+} directly activates regulatory enzymes, such as protein kinase C, which trigger signal transduction pathways.

The protein of SEQ ID NOS 301 and 542 is a calcium binding S100 protein typically found in heart and muscle. S100 proteins are low-molecular weight calcium binding proteins that are believed to play an important role in various cellular processes such as cytoskeletal/membrane interactions, cell division and differentiation. The expression of S100 proteins has been evaluated in a variety of disorders. For example, S100 protein have been evaluated as markers of inflammatory disease, including ulcerative colitis, Crohn's disease, and as serum markers for subjects with infections diseases including AIDS and malaria and for subjects with hematological disease. Evidence has accumulated that indicates that S100 proteins can alter cellular invasion and metastatic spread of cancer. For example, S100 protein is expressed in dendritic cells in human transitional cell carcinoma of the bladder and the invasive potential of these tumor has been found to correlate with the presence of S100 protein expressing cells.

THERAPEUTICS AND DIAGNOSTICS

The S100G protein of SEQ ID NOS 301 and 542 disclosed herein provides new calcium binding S100 protein compositions useful in the diagnosis, prevention, and treatment of cancer, reproductive disorders, immune disorders, neuronal disorders, vesicle trafficking disorders and developmental disorders.

CBPs are implicated in a variety of disorders and several CBPs have proven to be effective therapeutic targets for which small molecule inhibitors could be developed. However, while several CBPs are targets for widely-used therapeutic treatments, it would be advantageous to provide further CBPs allowing more selective therapeutic treatments for disease to be developed. In one example, calcineurin is found in the cells of all eukaryotes ranging from yeast to mammals. Calcineurin is a target for inhibition by the immunosuppressive agents cyclosporin A and FK506 emphasizing its importance in immune disorders (Kissinger, C. R. et al. (1995) Nature 378:641-644). Calcineurin also plays a critical role in transcriptional regulation and growth control in T-lymphocytes (Wang, M. G. et al. (1996) Cytogenet. Cell Genet. 72:236-241). However, inhibition

of calcineurin phosphatase activity has been implicated both in the mechanism of immunosuppression and in the observed toxic side effects of FK506 in nonlymphoid cells, suggesting that identification of a new (FK binding proteins (FKBPs) that can mediate calcineurin inhibition and are restricted in its expression to T cells could provide new immunosuppressive drugs

5 may be identified that, by virtue of their specific interaction with the FKBP, would be targeted in their site of action (Baughman G, et al Mol Cell Biol 1995 Aug;15(8):4395-402). In another CBP example, levels of CaM are increased several-fold in tumors and tumor-derived cell lines for various types of cancer (Rasmussen, C. D. and Means, A. R. (1989) Trends in Neuroscience 12:433-438). Calcium binding S100 β is another example of a CBP involved in a variety of

10 disorders. Like the S100G protein of the invention, S100 β contains an EF-hand motif. S100 β is abundantly expressed in the nervous system. S100 β levels are increased in the blood and cerebrospinal fluid of patients with neurological injury resulting from cerebral infarction, transient ischemic attacks, hemorrhagia, head trauma, and Down's syndrome. Furthermore, S100 β and other neural-specific CBPs may also protect against neurodegenerative disorders, such as Alzheimer's,

15 Parkinson's, and Huntington's diseases. S100 β is produced and secreted by glial cells in the central and peripheral nervous systems (Allore, R. J. et al. (1990) J. Biol. Chem. 265:15537-15543). The accumulation of S100 β in mature glial cells is associated with the microtubule network. S100 β promotes neuronal differentiation and survival but may be detrimental to cells if overexpressed. The selective overproduction has been implicated in the progression of the neuropathological changes in

20 Alzheimer's disease which may involve mitotic protein kinases (Marshak, D. R. and Pena, L. A. (1992) Prog. Clin. Biol. Res. 379:289-307). Adult T-cell leukaemia (ATL) is a mature T-cell malignancy which is caused by human T lymphotropic virus type-1. Diminished surface expression of the T-cell receptor alpha beta (TCR $\alpha\beta$ +) complex is a specific feature of ATL cells. S100 β is not detectable in CD4+, TCR $\alpha\beta$ +/ATL cells, but is expressed in CD4-, CD8-, TCR $\alpha\beta$ +

25 leukaemic cells from four ATL patients. This suggested that increased levels of S100 β may be associated with the diminished surface expression of the TCR $\alpha\beta$ +/ complex in ATL (Suzushima, H. et al. (1994) Leuk. Lymphoma 13:257-262). Elevated serum levels of S100 β are associated with disseminated malignant melanoma metastases, suggesting that serum S100 β may be of value as a clinical marker for progression of metastatic melanoma (Henze, G. et al. (1997) Dermatology

30 194:208-212). In yet another example, messenger RNA levels encoding human calgizzarin (an S100-like protein), as well as those encoding phospholipase A₂, are elevated in colorectal cancers compared with those of normal colorectal mucosa (Tanaka, M. et al. (1995) Cancer Lett. 89:195-200). Finally, an intracellular S100 calcium-binding protein has been isolated from rat peritoneum. This protein, MRP14, is one of two migration inhibitory factor-related proteins that are expressed in

35 peritoneal macrophages in the arthritis-susceptible Lewis/N rat (Imamichi, T. et al. (1993) Biochem. Biophys. Res. Comm. 194:819-825).

However, despite the many uses and therapeutics based on S100 proteins and other CBPs, it would be advantageous to selectively target CBPs which are involved in a disorder and which are found specifically in the targeted cells or tissues. Thus, in a first embodiment, the S100G protein of

the invention can be used for the development of selective inhibitors of calcium signaling, eg. preferably inhibitors of transcriptional regulation and cell growth control.

In one embodiment, an antagonist of S100G may be administered to a subject to prevent or treat a neuronal disorder. Such disorders may include, but are not limited to, akathisia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, schizophrenia, and Tourette's disorder. In one aspect, an antibody which specifically binds S100G may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express S100G.

In one embodiment, an antagonist of S100G may be administered to a subject to prevent or treat a vesicle trafficking disorder. Such disorders may include, but are not limited to, cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking including AIDS; allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminth, and protozoal infections. In one aspect, an antibody which specifically binds S100G may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express S100G.

In one embodiment, an antagonist of S100G may be administered to a subject to prevent or treat an immunological disorder. Such disorders may include, but are not limited to, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjogren's syndrome, Werner syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, and extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma. In one aspect, an antibody which specifically binds S100P may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express S100G.

In one embodiment, an antagonist of S100G may be administered to a subject to prevent or treat a neoplastic disorder. Such disorders may include, but are not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of

the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds S100P may be used directly as an antagonist or indirectly as a targeting or
5 delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express S100G.

In addition, the S100G nucleic acids and polypeptides of the present invention can be used to identify compounds for the treatment of a subject experiencing negative side effects from the administration of other pharmaceuticals, such as those drugs that disrupt the body's calcium homeostasis. Co-administration of said compounds would be useful to counter-effect iatrogenically
10 caused dysfunction of calcium metabolism.

An antagonist of S100G may be produced using methods which are generally known in the art. In particular, purified S100G may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind S100G. In one example, a positive screening for drugs that specifically inhibit the Ca²⁺-signaling activity was carried out on the basis
15 of the growth promoting effect on a yeast mutant with a peculiar phenotype (Shitamukai A et al, Biosci Biotechnol Biochem 2000 Sep;64(9):1942-6). An inappropriate activation of a signaling pathway in yeast often has a deleterious physiological effect and causes various defects, including growth defects. In a certain genetic background (*deltazds1*) of *Saccharomyces cerevisiae*, the cell-cycle progression in G2 is specifically blocked in the medium with CaCl₂ by the hyperactivation of
20 the Ca²⁺-signaling pathways. Shitamukai et al provide an example of a drug screening procedure designed to detect the active compounds that specifically attenuate the Ca²⁺-signaling activity on the basis of the ability to abrogate the growth defect of the cells suffering from the hyperactivated Ca²⁺ signal. Screening conditions were established for the drugs that suppress the Ca²⁺-induced growth inhibition using known calcineurin inhibitors as model compounds, and an indicator strain
25 with an increased drug sensitivity was constructed with a *syrl/erg3* null mutation.

In another embodiment, a vector expressing the complement of the polynucleotide encoding S100P may be administered to a subject to treat or prevent a neuronal disorder, immunological disorder, neoplastic disorder or vesicle trafficking disorder including, but not limited to, those described above. In other embodiments, any of the proteins, antagonists, antibodies, agonists,
30 complementary sequences or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to
35 achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Expression of S100 proteins have been evaluated as serum markers for melanoma (Henze et al, *Dermatology* 194:208-212; Buer et al, 1997 *Brit. J. Cancer* 75: 1373-1376; Sherbert et al. 1998; *Anticancer Res.* 18:2415-2422) and more recently as serum markers for cancer in general

(particularly breast, colon and lung cancers) and that their detectability in serum can have prognostic and/or therapeutic significance in cancer. Furthermore auto-antibodies to S100 proteins were found in cancer patients. (International Patent Publication No. WO 00/26668). The S100G protein of the invention may thus be used in the diagnosis and prevention of cancer, for

- 5 identification of subjects predisposed to cancer, for monitoring patients undergoing treatment for cancer based on the increased level of S100G protein(s) in biological fluid samples of subjects.

Methods for diagnosis and prognosis of cancer in a subject may comprise

- a) detecting a S100G protein in a biological fluid sample obtained from a subject, and
b) comparing the level of protein detected in the subject's sample to the level of protein
10 detected in a control sample,

wherein an increase in the level of S100G protein detected in the subject's sample as compared to control samples is an indicator of a subject with cancer or at increased risk for cancer.

The invention also comprises methods for diagnosis and prognosis of a subject with cancer comprising:

- 15 a) contacting a serum sample derived from a subject with a sample containing S100G protein antigens under conditions such that a specific antigen-antibody complex binding can occur; and

b) detecting the presence of immunospecific binding of autoantibodies present in the subject's serum samples to the S100G protein;

- 20 wherein the presence of immunospecific binding of autoantibodies indicates the presence of cancer.

Assays for detection of S100G protein in a sample can be accomplished by any suitable method, including immunoassays where in S100G proteins are detected by their interaction with an S100G specific antibody. In addition, reagents other than antibodies such as for example

- 25 polypeptides that specifically bind S100G may be used.

- In yet further embodiments, the S100G protein of the invention may be useful for the development of specific anti-S100 antibodies which are specific for S100 proteins other than S100G. As several S100 proteins have been implicated in disease, it may be advantageous to develop a panel of S100 specific antibodies to characterize disease, eg. cancers. Thus, S100G
30 proteins and antibodies thereto may be advantageous to distinguish cancer types. S100G proteins and antibodies may also be useful in the screening of S100 specific antibodies by determining the selectivity of a given anti-S100 antibody for its target, and eliminating antibodies which are cross-reactive with S100G proteins.

STRUCTURAL ASPECTS OF S100 PROTEINS OF THE INVENTION

- 35 The S100 proteins are a group of low molecular mass (approximately 10-12 kDa) acidic Ca^{2+} -binding proteins, so named after the solubility of the first isolated protein in 100% saturated ammonium sulfate. The most striking conserved feature of these proteins is the presence of an EF-hand. The S100 proteins have two Ca^{2+} -binding domains. One of these domains is a basic helix-

loop-helix domain, the other domain is an acidic helix-loop-helix EF-hand (Kligman, D. and Hilt, D. C. (1988) Trends Biochem. Sci. 13:437-442). The EF-hand domain also encompasses a part of a region within S100 proteins which specifically identifies members of the S100 family of proteins which have a low affinity for Ca^{2+} ions (S100/ICaBP; PROSITE PS00303, SWISSPROT, PFAM PF01023). The EF-hand is characterized by a twelve amino acid residue-containing loop, flanked by two alpha-helices, orientated approximately 90 degrees with respect to one another. Aspartate (D) and glutamate (E) residues are usually found bordering the twelve amino acid loop. In addition, a conserved glycine residue in the central portion of the loop is found in most Ca^{2+} -binding EF-hand domains. Oxygen ligands within this domain coordinate the Ca^{2+} ion (Kretsinger, R. H. and Nockolds, C. E. (1973) J. Biol. Chem. 248:3313-3326). It will also be appreciated by the skilled artisan that modifications of the S100G polypeptide may readily be made based on extensive knowledge of CBP structure and Ca^{2+} binding mechanisms, such as Sastry M et al, (Structure 1998;6:223-231), describing the three-dimensional structure of Ca^{2+} -bound calcyclin and its implications for Ca^{2+} -signal transduction by S100

15 Recently, a protein designated S100A13 has been discovered and characterized which shared significant primary structure similarity with the S100G polypeptide of the invention. (Wicki et al, (1996) BBRC 227:594-599).

The binding to calcium induces a conformational change in the S100 proteins, and this may then affect the secondary effector proteins. This mode of protein-protein interaction and modulation of the activity of the secondary effector protein is similar to that seen with calmodulin, also containing the EF-hands. The S100G protein of the invention comprises an ICaBP type calcium binding domain domain at amino acid positions 9 to 52 and casein kinase II phosphorylation site patterns at amino acid position 7 (TELE); 34 (SVNE); and 55 (SLDE) of SEQ ID NO 542.

The S100G polypeptides of the invention may thus also be used in any situation in vivo or in vitro where it is desired to modulate, preferably decrease the level of free calcium, or in applications involving sensing a change in calcium concentration. In one aspect, S100G nucleic acids and polypeptides may be used to develop a calcium biosensor. Calcium biosensors may have particular utility in detecting abnormalities in calcium transport that result in uncompensated influx into, or efflux from, the extracellular fluid, will result in hypercalcaemia or hypocalcaemia, respectively. Such abnormalities in serum calcium concentration may have profound effects on neurological, gastrointestinal, and renal function (Bushinsky DA et al, Lancet 1998 Jul 25;352 (9124):306-11). Calcium biosensors may be developed by using any suitable means known in the art to detect a conformational change induced by the binding of calcium to the EF-hand domains of the S100G polypeptides. Preferably, the conformational change is detected by detecting a change in the ability of the S100G protein to bind a selected secondary effector protein

As the distribution of particular S100 proteins is dependent on specific cell types, the S100 proteins may be involved in transducing the signal of an increase in intracellular calcium in a cell type-specific fashion (Wu, T. et al. (1997) J. Biol. Chem. 272:17145-17153).

CaMLP (CALCIUM BINDING PROTEIN) (Clone ID:500742698)

Calcium is one of the "second messengers" which relays chemical and electrical signals within a cell. This signal transduction and, hence the regulation of biological processes, involves interaction of calcium ion with high-affinity calcium-binding proteins (CBPs). Disclosed herein in

5 SEQ ID NOS 184 and 469 is one such protein, encoded by the nucleic acid sequences of SEQ ID NOS 15 and 352, respectively, and the clone FL11:500742698_204-61-4-0-B2-F, and further referred to herein as CaMLP, which is thought to act as a Ca^{2+} sensing and binding protein involved in diverse aspects of cell proliferation (such as for example of hepatocytes, melanoma cells, leukemic lymphocytes, and HUVEC (human umbilical vein endothelial cells)) and

10 differentiation. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOS:15 and 352 and polypeptides of SEQ ID NO:184 and 469, described throughout the present application also pertain to the human cDNA of clone 500742698, and the polypeptides encoded thereby. Notably, the CaMLP polypeptide contains EF hand calcium-binding domains (PROSITE PS00018) at amino acid positions 81-93 and at position 129-141 of SEQ ID NO 469. CaMLP

15 The cellular processes in which Ca^{2+} functions as an intracellular signaling molecule are diverse, including cell proliferation and differentiation, neurotransmitter secretion, glycogen metabolism, and skeletal muscle contraction. Within a resting cell, the concentration of Ca^{2+} in the cytosol is extremely low, $<10^{-7}$ M. However, when the cell is stimulated by an external signal, such as a neural impulse or a growth factor, the cytosolic concentration of Ca^{2+} increases by about

20 50-fold. This influx of Ca^{2+} is caused by the opening of plasma membrane Ca^{2+} channels and the release of Ca^{2+} from intracellular stores such as the endoplasmic reticulum. Ca^{2+} directly activates regulatory enzymes, such as protein kinase C, which trigger signal transduction pathways. Ca^{2+} also binds to specific Ca^{2+} -binding proteins (CBPs) such as calbindins, troponin C, calmodulin, and S-100 proteins which then activate multiple target proteins including enzymes, membrane

25 transport pumps, and ion channels. Calmodulin (CaM) is the most widely distributed and the most common mediator of calcium effects and appears to be the primary sensor of Ca^{2+} changes in eukaryotic cells. The binding of Ca^{2+} to CaM induces marked conformational changes in the protein permitting interaction with, and regulation of over 100 different proteins. CBP interactions are involved in a multitude of cellular processes including, but not limited to, gene regulation, DNA

30 synthesis, cell cycle progression, mitosis, cytokinesis, cytoskeletal organization, muscle contraction, signal transduction, ion homeostasis, exocytosis, and metabolic regulation (Celio, M. R. et al. (1996) Guidebook to Calcium-binding Proteins, Oxford University Press, Oxford, UK, pp. 15-20).

THERAPEUTICS

The CaMLP protein of SEQ ID NOS 184 and 469 disclosed herein provides new calcium

35 binding protein compositions useful in the diagnosis, prevention, and treatment of cancer, reproductive disorders, immune disorders, neuronal disorders and developmental disorders.

Calcium binding proteins (CBPs) are implicated in a variety of disorders and several CBPs have proven to be effective therapeutic targets for which small molecule inhibitors could be

developed. However, while several CBPs are targets for widely-used therapeutic treatments, it would be advantageous to provide further CBPs allowing more selective therapeutic treatments for disease to be developed. Evidence has accumulated for a large number of CBPs suggesting involvement in cell proliferative disorders. It is proposed that CaMLP may be useful as a tissue specific calmodulin homologue allowing the development of specific inhibitors and activators having increased selectivity and safety (decreased side effect profile). To date, calmodulin antagonists are reportedly useful for the treatment of some malignant tumors, particularly those of the central nervous system, as well as lung tumors. The antitumor activity of calmodulin antagonists, as well as successful chemotherapy using the same, has been described, for example, in Sculler et al. Cancer Res., 50:1645-1649 (1990) and Hait et al. Cancer Res., 50:6636-6640 (1990). U.S. Pat. No. 5,340,565, additionally describes the use of calmodulin antagonists or inhibitors as agents which enhance the effectiveness of a chemotherapeutic agent or radiation treatment. Specifically, described therein is a method of inhibiting or killing a tumor or cancer cell in a human patient undergoing radiation therapy or chemotherapy, for example with such chemotherapeutic agents as cisplatin (Platinol®), by additionally administering a calmodulin binding agent which inhibits calmodulin activity.

Calmodulin is also believed to play a pathogenic role in the tissue damage caused by burns and frostbite (Beitner et al., Gen. Pharmac. 20: 641-646, 1989), as well as in dermatitis and other conditions involving keratinocyte hyperproliferation. The methods of the present invention may be applied to the treatment of these and other conditions wherein antagonism of calmodulin activity is desirable.

Thus, as discussed above, CaMLP protein of the invention shares structural similarity with the ubiquitous intracellular receptor protein calmodulin suggesting that CaMLP may be useful in the development of selective CaMLP inhibitors. The nucleic acids and polypeptides of the invention thus provide a novel therapeutic target in particular for cell proliferative disorders. In one aspect, said nucleic acids and protein may be used in drug screening processes to develop selective calmodulin and other calcium binding protein antagonists which do not inhibit the polypeptide of the invention. In another aspect, the nucleic acids and polypeptides of the invention may be used in drug screening processes to identify selective modulators of the CaMLP without inhibiting calmodulin, thereby identifying compounds less likely to cause unwanted side effects. In yet another aspect, the nucleic acids and polypeptides of the invention may be used in drug screening processes to identify selective modulators of both calmodulin and CaMLP, thereby identifying compounds having increased potency.

Upon calcium binding, CaMLP may interact with a number of protein targets in a calcium dependent manner, thereby altering a number of complex biochemical pathways that can affect the overall behavior of cells. The calcium-calmodulin complex for example controls the biological activity of more than thirty different proteins including several enzymes, ion transporters, receptors, motor proteins, transcription factors, and cytoskeletal components in eukaryotic cells.

As described in U.S. Patent No. 5,840,697, Blondelle et al have peptide inhibitors of calmodulin. A number of other calmodulin targeted compounds are known and used for a variety of therapeutic applications. For instance, chlorpromazine (Thorazine.RTM.) and related phenothiazine derivatives, disclosed, for example, in U.S. Pat. No. 2,645,640, are calmodulin antagonists useful as
5 tranquilizers and sedatives. Naphthalenen-sulfonamides, also calmodulin antagonists, are known to inhibit cell proliferation, as disclosed, for example, in Hidaka et al. ((1981), PNAS, 78:4354-4357) and are useful as antitumor agents. In addition, the cyclic peptide cyclosporin A (Sandimmune®), disclosed in U.S. Patent No. 4,117,118, is as an immunosuppressive agent which is thought to work by inhibiting calmodulin mediated responses in lymphoid cells.

10 Many existing calmodulin inhibitors have undesirable biological effects when administered at concentrations sufficient to block calmodulin. These undesirable biological effects include non-specific binding to other proteins or receptors, as described, for example, in Polak et al, ((1991), J. Neurosci. 11:534-542.) In addition, negative side effects such as toxicity can occur. A specific example is the toxic side effects from cyclosporin A. Therefore, a need exists for calmodulin
15 targeted agents, and in particular calmodulin antagonists which inhibit calmodulin without having additional, undesirable biological or side effects. In particular there is a need for inhibitors which are specific to calmodulin and which do not have toxic side effects.

In addition, the CaMLP nucleic acids and polypeptides of the present invention can be used to identify compounds for the treatment of a subject experiencing negative side effects from the
20 administration of other pharmaceuticals, such as those drugs that disrupt the body's calcium homeostasis. Co-administration of said compounds would be useful to counter-effect iatrogenically caused dysfunction of calcium metabolism. Such disorders include, but are not limited to, organ damage, autoimmune disorders, psychotic disorders, tumors and drug induced dysfunction, such as negative side effects subsequent to administration of pharmaceuticals. For example, organ or tissue
25 transplantation can result in autoimmune disorders, such as tissue graft (allograft) rejections.

It is well known that calmodulin-targeted compounds which are antagonists can be used as immunosuppressive agents. In addition, also as described above, such compounds are widely used as sedative or anti-psychotic agents. Furthermore, there is evidence that calmodulin (and hence CaMLP) antagonists are useful for the treatment of some malignant tumors, particularly those of the
30 central nervous system, as well as lung tumors. The antitumor activity of calmodulin antagonists, as well as successful chemotherapy using the same, has been described, for example, in Sculler et al. Cancer Res., 50:1645-1649 (1990) and Hait et al. Cancer Res., 50:6636-6640 (1990), both of which are incorporated herein by reference. U.S. Pat. No. 5,340,565, which is incorporated herein by reference, additionally describes the use of calmodulin antagonists or inhibitors as agents which
35 enhance the effectiveness of a chemotherapeutic agent or radiation treatment. Specifically, described therein is a method of inhibiting or killing a tumor or cancer cell in a human patient undergoing radiation therapy or chemotherapy, for example with such chemotherapeutic agents as cisplatin (Platinol), by additionally administering a calmodulin binding agent which inhibits calmodulin activity.

It has also been found that extracellular calmodulin inhibits TNF release and facilitates elastase release, providing further suggestion that CaMLP, CaMLP analogues and CaMLP receptor agonists are useful agents for regulating the inflammatory process. CaMLP antagonists, which include CaMLP receptor antagonists and CaMLP -binding molecules, may be used to block the interaction of CaMLP with a receptor, thus providing the opposite effect from CaMLP, its analogues and receptor agonists. CaMLP may serve as a potent modulator of self-directed inflammation by assisting in the recognition of self vs. non-self as prokaryotes (e.g., bacterial pathogens) do not contain CaMLP. In some situations such as in tumor necrosis, release of extracellular CaMLP may lead to an inappropriate host response and failure of the immune/inflammatory systems to eradicate tumor cells. Further, a diagnostic test has been developed which can discern patient variabilities in TNF inhibition by calmodulin and other substances. This test can be utilized in monitoring individual patients for determining effective therapies, and for predicting efficacy of therapy with extracellular CaMLP, CaMLP analogues or CaMLP receptor agonists on the one hand and CaMLP antagonists on the other. A diagnostic test for elastase has also been developed with similar utility.

Lysozyme C Protein of SEQ ID NO: 196 (internal designation 482181) and related protein of SEQ ID NO:479

The polypeptides of SEQ ID NO : 196 and SEQ ID NO:479 encoded by the cDNA of SEQ ID NO:27 and 362, respectively, belong to the widely conserved family of lysozyme C precursors (Prager and Jollès, *Lysozymes: model enzymes in biochemistry and biology*, ed. Jollès, 9-321 (1996), Qasba and Kumar, *Crit. Rev. Biochem. Mol. Biol.* 32:255-306 (1997)), which disclosures are hereby incorporated by reference in their entireties. The protein of SEQ ID NOs:196 and 479 or part thereof plays a role in glycoprotein and/or peptidoglycan metabolism, probably as a glycosyl hydrolase of family 22. Thus, the protein of the invention or part thereof is involved in immune and inflammatory responses and has antiviral, antibacterial, anti-inflammatory and/or anti-histaminic functions. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:196 from positions 19 to 100, or from positions 1 to 100. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 196 having any of the biological activities described herein. The glycolytic activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in Gold and Schweiger, M. *Methods in Enzymology*, Vol. XX, Part C pp. 537-542, Ed. Moldave, Academic Press, New York and London, 1971 and in the US patent 4,255,517, which disclosures are hereby incorporated by reference in their entireties.

Lysozymes, which are ubiquitous proteins found in most body secretions, are defined as 1,4-beta-N-acetylmuramidase which cleave the glycoside bond between the C-1 of N-acetyl-muramic acid and the C-4 of N-acetylglucosamine in the peptidoglycan of bacteria. They have various therapeutic properties, such as antiviral, antibacterial, anti-inflammatory and antihistaminic

effects. The activity of lysozymes as an anti-bacterial agent appears to be based on both its direct bacteriolytic activity and also on stimulatory effects in connection with phagocytosis of polymorphonuclear leucocytes and macrophages (Biggar and Sturgess, J. M. Infect Immunol. 16: 974-982 (1977); Thacore and Willet, Am. Rev. Resp. Dis. 93: 786-790 (1966); Klockars and Roberts, P. Acta Haematol 55: 289-292 (1976)), which disclosures are hereby incorporated by reference in their entireties. Lysozymes have proven to be not only a selective factor but also an effective factor against microorganisms of the mouth (Iacono et al, J. J. Infect. Immunol. 29: 623-632 (1980)), which disclosure is hereby incorporated by reference in its entirety. Lysozymes can also kill pathogens by acting synergistically with other proteins such as complement or antibody to lyse pathogenic cells. Lysozymes, also inhibit chemotaxis of polymorphonuclear leukocytes and limit the production of oxygen free radicals following an infection. This limits the degree of inflammation, while at the same time enhances phagocytosis by these cells. Other postulated functions of lysozymes include immune stimulation (Jolles, P. Biomedicine 25: 275-276 (1976) Ossermann, E. F. Adv. Pathobiol 4: 98-102 (1976)) and immunological and non-immunological monitoring of host membranes for any neoplastic transformation (Jolles, P. Biomedicine 25: 275-276 (1976); Ossermann, E. F. Adv. Pathobiol 4: 98-102 (1976)), which disclosures are hereby incorporated by reference in their entireties. Lysozymes may thus be used in a wide spectrum of applications (see US patent 5,618,712, which disclosure is hereby incorporated by reference in its entirety). Determination of the lysozymes from serum and/or urine is used to diagnose various diseases or as an indicator for their development. In acute lymphoblastic leukaemia the lysozyme serum level is significantly reduced, whereas in chronic myelotic leukaemia and in acute monoblastic and myelomonocytic leukaemia the lysozyme concentration in the serum is greatly increased. The therapeutically effective use of lysozyme is possible in the treatment of various bacterial and virus infections (Zona, Herpes zoster), in colitis, various types of pain, in allergies, inflammation and in pediatrics (the conversion of cows milk into a form suitable for infants by the addition of lysozyme).

The invention relates to methods and compositions using the protein of the invention or part thereof to hydrolyze one or several substrates, alone or in combination with other substances, preferably antiviral, antifungal and/or antibacterial substances including but not limited to immunoglobulins, lactoferrin, betalysin, fibronectin, and complement components. Such substrates are glycosylated compounds, preferably containing beta-1-4-glycoside bonds, more preferably containing beta-1-4-glycoside bonds between n-acetylmuraminic acid and n-acetylglucosamine. For example, the protein of the invention or part thereof is added to a sample containing the substrate(s) in conditions allowing hydrolysis, and allowed to catalyze the hydrolysis of the substrate(s). In a preferred embodiment, the hydrolysis is carried out using a standard assay such as those described by Gold and Schweiger, supra, and US patents 5,871,477 and 4,255,517, which disclosures are hereby incorporated by reference in their entireties. In a preferred embodiment, the protein of the invention or part thereof may be used to lyse recombinant bacteria in order to recover the recombinant DNA, the recombinant protein of interest, or both using, for example, any of the

assays described in Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press (1989), which disclosure is hereby incorporated by reference in its entirety.

In an embodiment, the protein of the invention or part thereof is used to hydrolyze
5 contaminating substrates, preferably exogenous substrates from bacterial, fungal or viral origins, in an aqueous sample or onto a material, preferably glassware and plasticware. In particular, the protein of the invention or part thereof may be used as a disinfectant in dental rinse, in protection of aqueous systems or in preparing material for medical applications using any of the methods and compositions described in US patents 5,069,717, 4,355,022 and 5,001,062, which disclosures are
10 hereby incorporated by reference in their entireties. In a preferred embodiment, the protein of the invention is used as a host resistance factor in infants' formulas to convert cow's milk into a form more suitable for infants as described in US patent 6,020,015, which disclosure is hereby incorporated by reference in its entirety. In another preferred embodiment, the protein of the invention or part thereof may be used as a food preservative (see Hayashi et al., *Agric. Biol. Chem.*
15 (European Edition of Japanese Journal of Agriculture, Biochemistry and Chemistry), Vol. 53, pp. 3173-3177, 1989), which disclosure is hereby incorporated by reference in its entirety. In addition, the protein of the invention or part thereof may be used to clarify xanthan gum fermented broth for applications in food and in cosmetic industries using the method described in US patent 5,994,107, which disclosure is hereby incorporated by reference in its entirety. In another preferred
20 embodiment, compositions comprising the protein of the present invention or part thereof are added to samples or materials as a "cocktail" with other antimicrobial substances, preferably antibiotics or hydrolytic enzymes such as those described in US patents 5,458,876 and 5,041,326, which disclosures are hereby incorporated by reference in their entireties, to decontaminate the samples. For example, the protein of the invention or part thereof may be used in place or in combination
25 with antibiotics in cell cultures. The advantage of using a cocktail of hydrolytic enzymes is that one is able to hydrolyze a wide range of substrates without knowing the specificity of any of the enzymes. Using a cocktail of hydrolytic enzymes also protects a sample or material from a wide range of future unknown contaminants from a vast number of sources. For example, the protein of the invention or part thereof is added to samples where contaminating substrates, preferably
30 exogenous substrates from bacterial, fungal or viral origins, is undesirable in an amount sufficient to promote hydrolysis of said substrates. Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other hydrolytic enzymes, using techniques well known in the art, to form an affinity chromatography column. A sample containing the undesirable substrate is run through the column to remove the substrate.
35 Immobilizing the protein of the invention or part thereof on a support advantageous is particularly for those embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of the enzyme from the batch of product and subsequent reuse of the enzyme. Immobilization of the protein of the invention or part thereof can be accomplished, for example, by inserting a cellulose-binding domain in the protein. One of skill in

the art will understand that other methods of immobilization could also be used and are described in the available literature. Alternatively, the same methods may be used to identify new substrates.

In addition, the protein of the invention or part thereof may be useful to identify or quantify the amount of a given substrate in biological fluids, foods, water, air, solutions and the like. In a preferred embodiment, the protein of the invention or part thereof is used in assays and diagnostic kits for the identification and quantification of exogenous substrates in bodily fluids including blood, lymph, saliva or other tissue samples, in addition to bacterial, fungal, plant, yeast, viral or mammalian cell cultures. In a preferred embodiment, the protein of the invention or part thereof is used to detect, identify, and or quantify eubacteria using reagents and assays described in US patent 5,935,804, which disclosure is hereby incorporated by reference in its entirety. Briefly, the protein of the invention or part thereof is catalytically inactivated, i.e. capable of binding but not cleaving a peptidoglycan comprising NAc-muramic acid in the eubacteria, using any of the methods known to those skilled in the art including those which produce a mutant enzyme, a recombinant-enzyme, or a chemically inactivated enzyme. The catalytically inactive protein of the invention is then incubated with an aliquot of a biological sample under conditions suitable for binding of the inactive enzyme to the peptidoglycan substrate. Then, the bound enzyme is detected to assess the presence or amount of the eubacteria in the biological sample.

In another embodiment, the nucleic acid of the invention or part thereof may be used to increase disease resistance of plants to bacterial, fungal and/or viral infections. A polynucleotide containing the nucleic acid of the invention or part thereof is introduced into the plant genome in conditions allowing correct expression of the transgenic protein using any methods known to those skilled in the art including those disclosed in US patents 5,349,122 and 5,850,025, which disclosures are hereby incorporated by reference in their entireties.

In another preferred embodiment, the protein of the invention or part thereof may be useful to treat and/or prevent bacterial, fungal and viral infections in humans or in animals caused by various agents including but not limited to Streptococcus, Veillonella alcalescens, Actinomyces, Herpes simplex, Candida albicans, Micrococcus lysodeikticus and HIV by hydrolyzing the glycosylated compounds contained in such micro-organisms. In still a preferred embodiment, the protein of the invention or part thereof is used to prevent and/or treat bacterial, fungal and viral infections in immunocompromised individuals who lack fully functional immune systems, such as neonates or geriatric patients or HIV-infected individuals, or who suffer from a disease affecting the respiratory tract such as cystic fibrosis or the gastrointestinal tract such as ulcerative colitis or sprue.

In still another embodiment, the protein of the invention or part thereof may be used as a growth factor for in vitro cell culture, preferably for T cells and T cell lines, using techniques and methods taught in US patent 5,468,635, which disclosure is hereby incorporated by reference in its entirety.

In addition, the protein of the invention or part thereof may be used to identify inhibitors for mechanistic and clinical applications. Such inhibitors may then be used to identify or quantify the protein of the invention in a sample, and to diagnose, treat or prevent any of the disorders where the

protein's hydrolytic, immunostimulatory and/or inflammatory activities is/are undesirable and/or deleterious including but not limited to amyloidosis, colitis, lysosomal diseases, inflammatory and immune disorders including allergies and leukaemia. The protein of the invention may also be used to monitor host cell membranes for neoplastic transformation.

5 It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:27 and 362 and polypeptides of SEQ ID NO:196 and 479, described throughout the present application also pertain to the human cDNA of clone 482181, and the polypeptides encoded thereby.

10 Angiogenin Protein of SEQ ID NO: 176 (internal designation 114180) and related protein of SEQ ID NO:461

The polypeptides of SEQ ID NO:176 and SEQ ID NO:461 encoded by the extended cDNA SEQ ID NO:7 and SEQ ID NO:344, respectively, are ribonucleases that belongs to the pancreatic ribonuclease family (see reviews from Beintema, (1998) *Cell. Mol. Life Sci.* 54:763-5; Beintema and Kleineidam, (1998) *Cell. Mol. Life Sci.* 54:825-32, which disclosures are hereby incorporated by reference in their entireties). It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NO:7 and SEQ ID NO:344 and polypeptides of SEQ ID NO:176 and SEQ ID NO:461, described throughout the present application also pertain to the human cDNA of clone 114180, and the polypeptides encoded thereby. In addition, the protein of the invention plays a role in angiogenesis as an angiogenin variant protein (see review from Badet, (1999) *Pathol. Biol.* 74:345-51, which disclosure is hereby incorporated by reference in its entirety). Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO: 176 from positions 19 to 75, from positions 26 to 75, or from positions 63 to 69. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 176 having any of the biological activity described herein. The ribonuclease activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in US patent 5,866,119, which disclosure is hereby incorporated by reference in its entirety. The angiogenic activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described by Fett *et al.* (1985) *Biochem.* 24, 5480-5486, which disclosure is hereby incorporated by reference in its entirety.

Ribonucleases are proteins which catalyze the hydrolysis of phosphodiester bonds in RNA chains. Pancreatic ribonucleases are pyrimidic-specific ribonucleases present in high quantity in the pancreas of a number of mammalia taxa and of a few reptiles. In addition to their function in hydrolysis of RNA, ribonucleases have evolved to support a variety of other physiological activities. Such activities include anti-parasite, anti-bacterium, anti-virus, anti-neoplastic activities, neurotoxicity, and angiogenesis. For example, bovine seminal ribonuclease is anti-neoplastic (Laceetti, et al. (1992) *Cancer Res.* 52: 4582-4586, which disclosure is hereby incorporated by reference in its entirety). Some frog ribonucleases display both anti-viral and anti-neoplastic activity

(Youle, et al. (1994) *Proc. Natl. Acad. Sci. USA* 91: 6012-6016; Mikulski, et al. (1990) *J. Natl. Cancer Inst.* 82: 151-152; and Wu, et al. (1993) *J. Biol. Chem.* 268: 10686-10693), which disclosures are hereby incorporated by reference in their entireties. Eosinophil-derived neurotoxin (EDN) and eosinophil cationic protein (ECP) are related ribonucleases which possess neurotoxicity (Beintema, et al. (1988) *Biochemistry* 27: 4530-4538; Ackerman, (1993) In Makino, and Fukuda, *Eosinophils: Biological and Clinical Aspects*. CRC Press, Boca Raton, Fla., pp 33-74), which disclosures are hereby incorporated by reference in their entireties. In addition, ECP exhibits cytotoxic, anti-parasitic, and anti-bacterial activities. A EDN-related ribonuclease, named RNase k6, is shown to express in normal human monocytes and neutrophils, suggesting a role for this ribonuclease in host defense (Rosenberg, and Dyer, (1996) *Nuc. Acid. Res.* 24: 3507-3513), which disclosure is hereby incorporated by reference in its entirety.

Angiogenin is a tRNA-specific ribonuclease which binds protein partners on the surface of endothelial cells for endocytosis. Potential partners of angiogenin include heparin, plasminogen, elastase, angiostatin, actin, and a 170 kDa receptor on the surface of endothelial cells [Strydom, (1998) *Cell. Mol. Life Sci.* 54, 811-824, which disclosure is hereby incorporated by reference in its entirety]. Endocytosed angiogenin is translocated to the nucleus where it promotes endothelial invasiveness required for blood vessel formation (Moroianu, and Riordan, (1994) *Proc. Natl. Acad. Sci. USA* 91: 1217-1221, which disclosure is hereby incorporated by reference in its entirety).

Although originally isolated from medium conditioned by human colon cancer cells (Fett et al. (1985), *supra*), and subsequently shown to be produced by several other histological types of human tumors [Rybak, et al. (1987) *Biochem. Biophys. Res. Commun.* 146, 1240-1248; Olson, et al., (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 442-446, which disclosures are hereby incorporated by reference in their entireties], angiogenin also is a constituent of human plasma and normally circulates at a concentration of 250-360 ng/ml [Shimoyama, et al. (1996) *Cancer Res.* 56, 2703-2706; Blaser, et al. (1993) *Eur. J. Clin. Chem. Clin. Biochem.* 31, 513-516, which disclosures are hereby incorporated by reference in their entireties]. It has also been shown that recurrent gastric cancer patients had a much higher serum concentration of angiogenin than primary gastric cancer patients [Shimoyama, and Kaminishi, (2000) *J. Cancer Res. Clin. Oncol.* 126, 468-474, which disclosure is hereby incorporated by reference in its entirety].

Angiogenin is a potent inducer of angiogenesis [Fett, et al. *supra*]. Angiogenesis is a complex process of blood vessel formation comprising of several separate but interconnected steps at the cellular and biochemical level including: (i) activation of endothelial cells by the action of an angiogenic stimulus, (ii) adhesion and invasion of activated endothelial cells into the surrounding tissues and migration toward the source of the angiogenic stimulus, and (iii) proliferation and differentiation of endothelial cells to form a new microvasculature [Folkman, and Shing, (1992) *J. Biol. Chem.* 267, 10931-10934; Moscatelli, and Rifkin, (1988) *Biochim. Biophys. Acta* 948, 67-85, which disclosures are hereby incorporated by reference in their entireties]. While angiogenesis is a tightly-controlled process under usual physiological conditions, abnormal angiogenesis can have devastating consequences in pathological conditions such as arthritis, diabetic retinopathy and

tumor growth. It is now well-established that the growth of virtually all solid tumors is angiogenesis dependent [Folkman, (1989) J. Natl. Cancer Inst. 82, 4-6, which disclosure is hereby incorporated by reference in its entirety]. Angiogenesis is also a prerequisite for the development of metastasis, since it provides the means whereby tumor cells disseminate from the original primary tumor and
5 establish at distant sites [Mahadevan, and Hart, (1990) Rev. Oncol. 3, 97-103; Blood, and Zetter (1990) Biochim. Biophys. Acta 1032, 89-118, which disclosures are hereby incorporated by reference in their entirety]. Therefore, interference with the process of tumor-induced angiogenesis can be an effective therapy for both primary and metastatic cancers. Indeed, several anti-angiogenic agents have been produced and are currently in the clinical trial stage.

10 The invention relates to methods and compositions using the protein of the invention or part thereof to hydrolyze one or several substrates, preferably nucleic acids, more preferably RNA, alone or in combination with other substances. For example, the protein of the invention or part thereof is added to a sample containing the substrate(s) in conditions allowing hydrolysis, and allowed to catalyze the hydrolysis of the substrate(s). Hydrolysis conditions as described in the US patent
15 5,866,119 may be used, which disclosure is hereby incorporated by reference in its entirety.

In a preferred embodiment, the protein of the invention or part thereof may be used to remove contaminating RNA in a biological sample, alone or in combination with other nucleases. In a more preferred embodiment, the protein of the invention or part thereof may be used to purify DNA preparations from contaminating RNA, to remove RNA templates prior to second strand
20 synthesis and prior to analysis of in vitro translation products. Compositions comprising the protein of the present invention or part thereof are added to biological samples as a "cocktail" with other nucleases. The advantage of using a cocktail of hydrolytic enzymes is that one is able to hydrolyze a wide range of substrates without knowing the specificity of any of the enzymes. Such cocktails of nucleases are commonly used in molecular biology assays, for example to remove unbound RNA in
25 RNase protection assays. Using a cocktail of hydrolytic enzymes also protects a sample from a wide range of future unknown RNA contaminants from a vast number of sources. For example, the protein of the invention or part thereof is added to samples where contaminating substrates is undesirable. Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other hydrolytic enzymes, using
30 techniques well known in the art, to form an affinity chromatography column. A sample containing the undesirable substrate is run through the column to remove the substrate. Immobilizing the protein of the invention or part thereof on a support is particularly advantageous for those embodiments in which the method is to be practiced on a commercial scale. This immobilization facilitates the removal of the enzyme from the batch of product and subsequent reuse of the enzyme.
35 Immobilization of the protein of the invention or part thereof can be accomplished, for example, by inserting a cellulose-binding domain in the protein. One of skill in the art will understand that other methods of immobilization could also be used and are described in the available literature. Alternatively, the same methods may be used to identify new substrates.

In another embodiment, the protein of the invention or part thereof may be used to decontaminate or disinfect samples infected by undesirable parasite, bacteria and/or viruses using any of the methods known to those skilled in the art including those described in Youle et al, (1994), supra; Mikulski et al (1990) supra, Wu et al (1993) supra.

5 In another embodiment, the present invention relates to compositions and methods using the protein of the invention or part thereof to selectively kill cells. The protein of the invention or part thereof is linked to a recognition moiety capable of binding to a chosen cell, such as lectins, receptors or antibodies thus generating cytotoxic reagents using methods and techniques described in US patent 5,955,073, which disclosure is hereby incorporated by reference in its entirety.

10 In still another embodiment, the invention relates to compositions and methods using the protein of the invention or part thereof to stimulate cell proliferation both in vitro and in vivo, especially endothelial cell growth. For example, soluble forms of the protein of the invention or part thereof may be added to cell culture medium in an amount effective to stimulate cell proliferation.

15 In still another embodiment, the protein of the invention or part thereof may be used in the diagnosis, prevention and/or treatment of disorders associated with excessive angiogenesis such as tumor growth, arthritis or diabetic retinopathy.

In a preferred embodiment, the protein of the invention may be used as a diagnostic marker to evaluate the risk of a given individual to develop a tumor, to evaluate the risk of recurrence of
20 tumors or to evaluate the degree of cancer aggressiveness based on the facts that the level of circulating angiogenin is lower in normal individuals than in patients bearing tumors, and is lower in patients with primary cancers compared to patients with reoccurrent tumors, as stated above. Thus, quantitative immunoassays can be used for the detection of abnormal levels of either the protein of SEQ ID NO: 176 or the mRNA encoding such protein as the polynucleotide of SEQ ID
25 No:7, thereby identifying those individuals at risk for the development of tumors or the recurrence of tumors. Detection of abnormal levels of the protein of the invention may be performed using any techniques known to those skilled in the art including those described elsewhere in the application. For example, antibodies binding specifically to the protein of the invention, or fragments thereof, may be used in routine immunoassays to screen for the presence or absence of the protein of the
30 invention, or fragments thereof. Alternatively, the nucleic acids which encode the protein of the invention, or fragments thereof, may be used in hybridization assays to detect and/or quantify the expression of said protein.

Another aspect of the invention provides for molecules which inhibit, or reduce, the biological activity or expression of SEQ ID NO: 276. Such molecules may be administered to
35 patients to prevent vascularization, especially tumor vascularization, thereby limiting tumor growth. Such antagonists and/or inhibitors may be antibodies specific for the protein of the invention that can be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein of the invention. Neutralizing antibodies, (i.e., those which inhibit protein-protein interactions) are especially preferred for

therapeutic use. Alternatively, such molecules may be mutated forms of the protein of the invention or truncated forms which will be able to bind to the partners of the protein of the invention and compete with it for partners but without eliciting any of its biological activities. Other methods to inhibit the expression of the protein of the invention include antisense and triple helix strategies as described herein. Other antagonists or inhibitors of the protein of the invention may be produced using methods which are generally known in the art, including the screening of libraries of pharmaceutical agents to identify those which specifically bind the protein of the invention. The protein of the invention, or part thereof, preferably its functional or immunogenic fragments, or oligopeptides related thereto, can be used for screening libraries of compounds in any of a variety of drug screening techniques including those described herein.

Protease Inhibitor Protein of SEQ ID NO: 181 (internal designation 1000771934) and related protein of SEQ ID NO:466

The protein of SEQ ID NOs:181 and 466 encoded by the extended cDNA SEQ ID NO:12 and 349, respectively, is a protease inhibitor belonging to the WAP-type disulfide core family. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:12 and 349 and polypeptides of SEQ ID NOs:181 and 466, described throughout the present application also pertain to the human cDNA of clone 1000771934, and the polypeptides encoded thereby. Preferred polypeptides of the invention are polypeptides comprising the amino acid fragments of SEQ ID NO:181 from positions 32 to 73, 49 to 62, 76 to 122, 97 to 110 or any combination thereof. Other preferred polypeptides of the invention are fragments of SEQ ID NO:181 having any of the biological activity described herein. The protease inhibitor activity of the protein of the invention or part thereof may be assessed using any techniques known to those skilled in the art. Possible substrates for the protein of the invention include are not limited to trypsin, chymotrypsin, leukoproteinase, elastase, subtilisin, type IV collagenase and other serine proteases.

Proteases, which cleave proteins, are largely used in industry including in food processing, brewing, and alcohol production. Proteases are important components of laundry detergents and other products. Within biological research, proteases are used in purification processes to degrade unwanted proteins. It is often desirable to employ proteases of low specificity or mixtures of more specific proteases to obtain the necessary degree of degradation.

Proteases are also key components of a broad range of biological pathways, including blood coagulation and digestion. For example, the absence or insufficiency of a protease can result in a pathological condition that can be treated by replacement or augmentation therapy. Such therapies include the treatment of hemophilia with clotting factors VIII, IX, and VIIa. In another application, the proteolytic enzyme tissue plasminogen activator (t-PA) is used to activate the body's clot lysing mechanism, thereby reducing morbidity resulting from myocardial infarction. The protease thrombin is used to initiate the clotting of fibrinogen-based tissue adhesives during surgery. Neutrophils produce several antibacterial serine proteases (Gabay, Ciba Found. Symp. 186:237-247, 1994;

Scocchi et al., Eur. J. Biochem. 209:589-595, 1992, which disclosures are hereby incorporated by reference in their entireties). Proteases also regulate cellular processes through receptor-mediated pathways by proteolytic activation of the cognate receptor (Vu et al., Cell 64:1057-1068, 1991; Blackhart et al., J. Biol. Chem. 271:16466-16471, 1996, which disclosures are hereby incorporated
5 by reference in their entireties).

Overproduction or lack of regulation of proteases can also have pathological consequences. Elastase, released within the lung in response to the presence of foreign particles, can damage lung tissue if its activity is not tightly regulated. Emphysema in smokers is believed to arise from an imbalance between elastase and its inhibitor, alpha-1-antitrypsin. This balance may be restored by
10 administration of exogenous alpha-1-antitrypsin.

In addition, protease inhibitors have been shown to inhibit the growth of microorganisms including human pathogenic bacteria such as strains of group A streptococci, including antibiotic-resistant strains (Merigan, T. et al (1996) Ann Intern Med 124:1039-1050; Stoka, V. (1995) FEBS. Lett 370:101-104; Vonderfecht, S. et al (1988) J Clin Invest 82:2011-2016; Collins, A. et al (1991)
15 Antimicrob Agents Chemother 35:2444-2446, which disclosures are hereby incorporated by reference in their entireties).

In view of the growing use of proteases in industry, research, and medicine, there is an ongoing need in the art for new enzymes and new enzyme inhibitors. The present invention addresses these needs.

20 The invention relates to compositions and methods using the protein of the invention or part thereof to inhibit proteases, both in vitro or in vivo. Since proteases play an important role in the regulation of many biological processes in virtually all living organisms as well as a major role in diseases, inhibitors of proteases are useful in a wide variety of applications.

In one embodiment, the protein of the invention or part thereof may be useful to quantify
25 the amount of a given protease in a biological sample, and thus used in assays and diagnostic kits for the quantification of proteases in bodily fluids or other tissue samples, in addition to bacterial, fungal, plant, yeast, viral or mammalian cell cultures. In a preferred embodiment, the sample is assayed using a standard protease substrate. A known concentration of protease inhibitor is added, and allowed to bind to a particular protease present. The protease assay is then rerun, and the loss
30 of activity is correlated to the protease inhibitor activity using techniques well known to those skilled in the art.

In addition, the protein of the invention or part thereof may be used to remove, identify or inhibit contaminating proteases in a sample. Compositions comprising the polypeptides of the present invention may be added to biological samples as a "cocktail" with other protease inhibitors
35 to prevent degradation of protein samples. The advantage of using a cocktail of protease inhibitors is that one is able to inhibit a wide range of proteases without knowing the specificity of any of the proteases. Using a cocktail of protease inhibitors also protects a protein sample from a wide range of future unknown proteases which may contaminate a protein sample from a vast number of sources. For example, the protein of the invention or part thereof are added to samples where

proteolytic degradation by contaminating proteases is undesirable. Such protease inhibitor cocktails (see for example the ready to use cocktails sold by Sigma) are widely used in research laboratory assays to inhibit proteases susceptible of degrading a protein of interest for which the assay is to be performed. Alternatively, the protein of the invention or part thereof may be bound to a
5 chromatographic support, either alone or in combination with other protease inhibitor, using techniques well known in the art, to form an affinity chromatography column. A sample containing the undesirable protease is run through the column to remove the protease. Alternatively, the same methods may be used to identify new proteases.

In a preferred embodiment, the protein of the invention or part thereof may be used to
10 inhibit proteases implicated in a number of diseases where cellular proteolysis occur such as diseases characterized by tissue degradation including but not limited to arthritis, muscular dystrophy, inflammation, tumor invasion, glomerulonephritis, parasite-borne infections, Alzheimer's disease, periodontal disease, and cancer metastasis.

In another preferred embodiment, the protein of the invention or part thereof may be useful
15 to inhibit exogenous proteases, both in vivo and in vitro, implicated in a number of infectious diseases including but not limited to gingivitis, malaria, leishmaniasis, filariasis, osteoporosis and osteoarthritis, and other bacterial, and parasite-borne or viral infections. In particular, the protein of the invention or part thereof may offer applications in viral diseases where the proteolysis of primary polypeptide precursors is essential to the replication of the virus, as for HIV and HCV.

20 In another embodiment, the protease inhibitors of the present invention may be used as antibacterial agents to retard or inhibit the growth of certain bacteria either in vitro or in vivo. Particularly, the polypeptides of the present invention may be used to inhibit the growth of group A streptococci on non-living matter such as surgical instruments, laboratory glassware and plasticware, and in culture of living plant, fungi, and animal cells.

25 Furthermore, the protease inhibitors of the present invention find use in drug potentiation applications. For example, therapeutic agents such as antibiotics or antitumor drugs can be inactivated through proteolysis by endogenous proteases, thus rendering the administered drug less effective or inactive. Accordingly, the protease inhibitors of the invention may be administered to a patient in conjunction with a therapeutic agent in order to potentiate or increase the activity of the
30 drug. This co-administration may be by simultaneous administration, such as a mixture of the protease inhibitor and the drug, or by separate simultaneous or sequential administration.

Serpin Protein of SEQ ID NO: 179 (internal designation 784093) and related protein of SEQ ID NO:464

35 The protein of SEQ ID NOs:179 and 464 and encoded by the extended cDNA SEQ ID NOs:10 and 347, respectively, is a serine protease inhibitor. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:10 and 347 and polypeptides of SEQ ID NOs:179 and 464, described throughout the present application also pertain to the human cDNA

of clone 784093, and the polypeptides encoded thereby. Preferred polypeptides of the invention are polypeptides comprising the amino acid fragments of SEQ ID NO:179 from positions 47 to 139. Other preferred polypeptides of the invention are fragments of SEQ ID NO:179 having any of the biological activity described herein. The protease inhibitor activity of the protein of the invention
5 or part thereof may be assessed using any techniques known to those skilled in the art including those disclosed in the US patent 5,955,284. Possible substrates for the protein of the invention include are not limited to serine proteases such as elastase, trypsin, chymotrypsin, thrombin III, plasmin, heparin, complement II, plasminogen activator, protein C, interleukin-1B converting enzyme, preferably trypsin, elastase, and chymotrypsin.

10 Proteases are key components of a broad range of biological pathways, including blood coagulation and digestion. For example, the absence or insufficiency of a protease can result in a pathological condition that can be treated by replacement or augmentation therapy. Such therapies include the treatment of hemophilia with clotting factors VIII, IX, and VIIa. In another application, the proteolytic enzyme tissue plasminogen activator (t-PA) is used to activate the body's clot lysing
15 mechanism, thereby reducing morbidity resulting from myocardial infarction. The protease thrombin is used to initiate the clotting of fibrinogen-based tissue adhesives during surgery. Neutrophils produce several antibacterial serine proteases (Gabay, Ciba Found. Symp. 186:237-247, 1994; Scocchi et al., Eur. J. Biochem. 209:589-595, 1992, which disclosures are hereby incorporated by reference in their entirety). Proteases also regulate cellular processes through receptor-mediated
20 pathways by proteolytic activation of the cognate receptor (Vu et al., Cell 64:1057-1068, 1991; Blackhart et al., J. Biol. Chem. 271:16466-16471, 1996, which disclosures are hereby incorporated by reference in their entirety).

Overproduction or lack of regulation of proteases can also have pathological consequences. Elastase, released within the lung in response to the presence of foreign particles, can damage lung
25 tissue if its activity is not tightly regulated. Emphysema in smokers is believed to arise from an imbalance between elastase and its inhibitor, alpha-1-antitrypsin. This balance may be restored by administration of exogenous alpha-1-antitrypsin.

The serine proteases (SP) are a large family of proteolytic enzymes that include the digestive enzymes, trypsin and chymotrypsin, components of the complement cascade and of the
30 blood-clotting cascade, and enzymes that control the degradation and turnover of macromolecules of the extracellular matrix. SP are so named because of the presence of a serine residue in the active catalytic site for protein cleavage. They are characterized by a catalytic triad of serine, histidine, and aspartic acid residues. SP have a wide range of substrate specificities and can be subdivided into subfamilies on the basis of these specificities. The main sub-families are trypases (cleavage after
35 arginine or lysine), aspases (cleavage after aspartate), chymases (cleavage after phenylalanine or leucine), metases (cleavage after methionine), and serases (cleavage after serine).

Serine proteases are used for a variety of industrial purposes. For example, the serine protease subtilisin is used in laundry detergents to aid in the removal of proteinaceous stains (e.g., Crabb, ACS Symposium Series 460:82-94, 1991, which disclosure is hereby incorporated by

reference in its entirety). In the food processing industry, serine proteases are used to produce protein-rich concentrates from fish and livestock, and in the preparation of dairy products (Kida et al., Journal of Fermentation and Bioengineering 80:478-484, 1995; Haard and Simpson, in Martin, A. M., ed., Fisheries Processing: Biotechnological Applications, Chapman and Hall, London, 1994, 5 132-154; Bos et al., European Patent Office Publication 494 149 A1, which disclosures are hereby incorporated by reference in their entireties).

Serpins are irreversible serine protease inhibitors which are principally located extracellularly. Proteins which have been assigned to the serpin family include the following: .alpha.-1 protease inhibitor, .alpha.-1-antichymotrypsin, antithrombin III, .alpha.-2-antiplasmin, 10 heparin cofactor II, complement C1 inhibitor, plasminogen activator inhibitors 1 and 2, glia derived nexin, protein C inhibitor, rat hepatocyte inhibitors, crmA (a viral serpin which inhibits interleukin 1-.beta. cleavage enzyme), human squamous cell carcinoma antigen which may modulate the host immune response against tumor cells, human maspin which seems to function as a tumor suppressor, lepidopteran protease inhibitor, leukocyte elastase inhibitor (the only known 15 intracellular serpin), and products from three orthopoxviruses (these products may be involved in the regulation of the blood clotting cascade and/or of the complement cascade in the mammalian host).

In view of the growing use of proteases in industry, research, and medicine, there is an ongoing need in the art for new enzymes and new enzyme inhibitors. The present invention 20 addresses these needs.

In one embodiment, the protein of the invention or part thereof may be useful to quantify the amount of a given protease in a biological sample, and thus used in assays and diagnostic kits for the quantification of proteases in bodily fluids or other tissue samples, in addition to bacterial, fungal, plant, yeast, viral or mammalian cell cultures. In a preferred embodiment, the sample is 25 assayed using a standard protease substrate. A known concentration of protease inhibitor is added, and allowed to bind to a particular protease present. The protease assay is then rerun, and the loss of activity is correlated to the protease inhibitor activity using techniques well known to those skilled in the art. Preferred proteases in this embodiment are serine protease, more preferably elastase, trypsin and chymotrypsin.

30 In addition, the protein of the invention or part thereof may be used to remove, identify or inhibit contaminating proteases in a sample. Compositions comprising the polypeptides of the present invention may be added to biological samples as a "cocktail" with other protease inhibitors to prevent degradation of protein samples. The advantage of using a cocktail of protease inhibitors is that one is able to inhibit a wide range of proteases without knowing the specificity of any of the 35 proteases. Using a cocktail of protease inhibitors also protects a protein sample from a wide range of future unknown proteases which may contaminate a protein sample from a vast number of sources. For example, the protein of the invention or part thereof are added to samples where proteolytic degradation by contaminating proteases is undesirable. Such protease inhibitor cocktails (see for example the ready to use cocktails sold by Sigma) are widely used in research laboratory

assays to inhibit proteases susceptible of degrading a protein of interest for which the assay is to be performed. Alternatively, the protein of the invention or part thereof may be bound to a chromatographic support, either alone or in combination with other protease inhibitor, using techniques well known in the art, to form an affinity chromatography column. A sample containing
5 the undesirable protease is run through the column to remove the protease. Alternatively, the same methods may be used to identify new proteases.

In a preferred embodiment, the protein of the invention or part thereof may be used to inhibit proteases implicated in a number of diseases where cellular proteolysis occur such as diseases characterized by tissue degradation including but not limited to arthritis, muscular
10 dystrophy, inflammation, tumor invasion, glomerulonephritis, parasite-borne infections, Alzheimer's disease, periodontal disease, and cancer metastasis. In a more preferred embodiment, the invention relates to compositions and methods to use the protein of the invention or part thereof in diseases characterized by an abnormally elevated levels of trypsin, chymotrypsin or elastase, including but not limited to chronic emphysema of the lungs, cirrhosis, liver failure, cystic fibrosis,
15 alpha1-antitrypsin deficiency associated disorders such as aneurysm or toxic shock. For prevention and/or treatment purposes, the protein of the invention may be used using any of the gene therapy methods described herein or known to those skilled in the art.

In another preferred embodiment, the protein of the invention or part thereof may be useful to inhibit exogenous proteases, both in vivo and in vitro, implicated in a number of infectious
20 diseases including but not limited to gingivitis, malaria, leishmaniasis, filariasis, osteoporosis and osteoarthritis, and other bacterial, and parasite-borne or viral infections. In particular, the protein of the invention or part thereof may offer applications in viral diseases where the proteolysis of primary polypeptide precursors is essential to the replication of the virus, as for HIV and HCV.

In another embodiment, the protease inhibitors of the present invention may be used as
25 antibacterial agents to retard or inhibit the growth of certain bacteria either in vitro or in vivo. Particularly, an amount of the polypeptides of the present invention effective to inhibit proliferation may be used to inhibit the growth of group A streptococci on non-living matter such as surgical instruments, laboratory glassware and plasticware, and in culture of living plant, fungi, and animal cells.

30 Furthermore, the protease inhibitors of the present invention find use in drug potentiation applications. For example, therapeutic agents such as antibiotics or antitumor drugs can be inactivated through proteolysis by endogenous proteases, thus rendering the administered drug less effective or inactive. Accordingly, the protease inhibitors of the invention may be administered to a patient in conjunction with a therapeutic agent in order to potentiate or increase the activity of the
35 drug. This co-administration may be by simultaneous administration, such as a mixture of the protease inhibitor and the drug, or by separate simultaneous or sequential administration.

ATPKf Protein sequence of SEQ ID No.253 (internal designation 1000867870)

The protein of SEQ ID NO:253 encoded by the extended cDNA SEQ ID NO:84 and relate polynucleotides of SEQ ID NO:404 is a variant of the human mitochondrial ATP synthase f subunit or ATPK (E.C. 3.6.1.34) and, as such, plays a role in cellular respiration. Preferred polypeptides of the invention are are polypeptides comprising the amino acids of SEQ ID NO: 253 from positions 5 to 88. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 253 having any of the biological activity described herein. It will be appreciated that all characteristics and uses of the polynucleotides of SEQ ID NOs:84 and 404 and polypeptides of SEQ ID NO: 253 described throughout the present application also pertain to the human cDNA of clone 1000867870, and the polypeptides encoded thereby.

The mitochondrial electron transport (or respiratory) chain is a series of enzyme complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP then provides the primary source of energy for driving a cell's many energy-requiring reactions.

ATP synthase (F₀ F₁ ATPase) is the enzyme complex at the terminus of this chain and serves as a reversible coupling device that interconverts the energies of an electrochemical proton gradient across the mitochondrial membrane into either the synthesis or hydrolysis of ATP. This gradient is produced by other enzymes of the respiratory chain in the course of electron transport from NADH to oxygen. When the cell's energy demands are high, electron transport from NADH to oxygen generates an electrochemical gradient across the mitochondrial membrane. Proton translocation from the outer to the inner side of the membrane drives the synthesis of ATP. Under conditions of low energy requirements and when there is an excess of ATP present, this electrochemical gradient is reversed and ATP synthase hydrolyzes ATP. The energy of hydrolysis is used to pump protons out of the mitochondrial matrix. ATP synthase is, therefore, a dual complex, the F₀ portion of which is a transmembrane proton carrier or pump, and the F₁ portion of which is catalytic and synthesizes or hydrolyzes ATP. Mammalian ATP synthase complex consists of sixteen different polypeptides (Walker, J. E. and Collinson, T. R. (1994) FEBS Lett.346: 39-43, which disclosure is hereby incorporated by reference in its entirety). Six of these polypeptides (subunits alpha, beta, gamma, delta, epsilon, and an ATPase inhibitor protein IF₁) comprise the globular catalytic F₁ ATPase portion of the complex, which lies outside of the mitochondrial membrane. The remaining ten polypeptides (subunits a, b, c, d, e, f, g, F₆, OSCP, and A6L) comprise the proton-translocating, membrane spanning F₀ portion of the complex. Like other members of the respiratory chain, all but two of the polypeptide subunits of ATP synthase are nuclear gene products that are imported into the mitochondria. Enzyme complexes similar to mammalian ATP synthase are found in all cell types and in chloroplast and bacterial membranes. This universality indicates the central importance of this enzyme to ATP metabolism. Transcriptional regulation of these nuclear encoded genes appears to be the predominant means for controlling the biogenesis of ATP synthase. Multiple mitochondrial pathologies exist because of the essential role of mitochondrial oxidative phosphorylation in cellular energy production, in the generation of reactive oxygen species and in

the initiation of apoptosis (Wallace, Science, 283:1482-1488, 1999, which disclosure is hereby incorporated by reference in its entirety). It is now clear that mitochondrial diseases encompass an assemblage of clinical problems commonly involving tissues that have high energy requirements such as heart, muscle and the renal and endocrine systems. Over the past 11 years, a considerable
5 body of evidence has accumulated implicating defects in the mitochondrial energy-generating pathway, oxidative phosphorylation, in a wide variety of degenerative diseases including myopathy and cardiomyopathy. Most classes of pathogenic mitochondrial DNA mutations affect the heart, in association with a variety of other clinical manifestations that can include skeletal muscle, the central nervous system (including eye), the endocrine system, and the renal system. Nuclear
10 mutations causing mitochondrial disorders have been described. They are often found in highly conserved subunits. Mitochondrial disorders with nuclear mutations include : myopathies (PEO, MNGIE, congenital muscular dystrophy, carnitine disorders), encephalopathies (Leigh, Infantile, Wilson's disease, Deafness-Dystonia syndrome), other systemic disorders and cardiomyopathies.

The discovery of a new ATP synthase subunit, and polynucleotides encoding it satisfy a
15 need in the art by providing new compositions which are useful for the diagnosis, prevention, and treatment of cancer, myopathies, immune disorders, and neurological disorders.

An object of the present invention relates to compositions and methods of targeting heterologous compounds, either polypeptides or polynucleotides to mitochondria by recombinantly or chemically fusing a fragment of the protein of the invention to an heterologous polypeptide or
20 polynucleotide. Preferred fragments are signal peptide, amphiphilic alpha helices and/or any other fragments of the protein of the invention, or part thereof, that may contain targeting signals for mitochondria including but not limited to matrix targeting signals as defined in Herrman and Neupert, Curr. Opin. Microbiol. 3:210-4 (2000); Bhagwat et al. J. Biol. Chem. 274:24014-22 (1999), Murphy Trends Biotechnol. 15:326-30 (1997); Glaser et al. Plant Mol Biol 38:311-38
25 (1998); Ciminale et al. Oncogene 18:4505-14 (1999), which disclosures are hereby incorporated by reference in their entireties. Such heterologous compounds may be used to modulate mitochondria's activities. For example, they may be used to induce and/or prevent mitochondrial-induced apoptosis or necrosis. In addition, heterologous polynucleotides may be used for mitochondrial gene therapy to replace a defective mitochondrial gene and/or to inhibit the
30 deleterious expression of a mitochondrial gene.

The invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which mitochondrial respiratory electron transport chain is impaired, including but not limited to mitochondriocytopathies, necrosis, aging, myopathies, cancer and neurodegenerative diseases such as Alzheimer's disease,
35 Huntington's disease, Parkinson's disease, epilepsy, Down's syndrome, dementia, multiple sclerosis, and amyotrophic lateral sclerosis. For diagnostic purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the protein of the invention may be used to enhance electron transport

and increase energy delivery using any of the gene therapy methods described herein or known to those skilled in the art.

In another embodiment, the invention further relates to methods and compositions using the protein of the invention or part thereof to diagnose, prevent and/or treat several disorders in which mitochondrial respiratory electron transport chain needs to be impaired, including but not limited to Sjogren's syndrome, Addison's disease, bronchitis, dermatomyositis, polymyositis, glomerulonephritis, diabetes mellitus, emphysema, Graves' disease, atrophic gastritis, lupus erythematosus, myasthenia gravis, multiple sclerosis, autoimmune thyroiditis, ulcerative colitis, anemia, pancreatitis, scleroderma, rheumatoid and osteoarthritis, asthma, allergic rhinitis, atopic dermatitis, dermatomyositis, polymyositis, and gout, using any techniques known to those skilled in the art including the antisense or triple helices strategies described herein.

Moreover, antibodies to the protein of the invention or part thereof may be used for detection of mitochondria organelles and/or mitochondrial membranes using any techniques known to those skilled in the art.

15

Oligomerization Protein sequence of SEQ ID No. 310 (internal designation D150568)

The protein of SEQ ID NO : 310 encoded by the cDNA of SEQ ID NO: 141 and 435, is able to form homo-oligomers. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:310 from positions 1 to 109. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 310 having any of the biological activities described herein.

Multivalency is a prerequisite for a variety of macromolecular interactions such as binding of antibodies or lectins to specific targets, ligand recognition, activation or inhibition of receptors and cell adhesion. Dimerization and oligomerization of proteins are thus general biological control mechanisms that contribute to the activation of cell membrane receptors, transcription factors, vesicle fusion proteins, and other classes of intra- and extracellular proteins.

Multimerization domains have been shown to be useful tools in several areas of biotechnology, especially in protein engineering. For example, Tso et al have used leucine zippers for producing bispecific antibody heterodimers (US patent 5,932,448) / Methods of preparing soluble oligomeric proteins using leucine zippers have been described by Conrad et al (US patent 5,965,712), Ciardelli et al (US patent 5,837,816), Spriggs et al (WO9410308) / Leucine zipper forming sequences have been used by Pelletier et al in protein fragment complementation assays to detect biomolecular interactions (WO9834120), which disclosures are hereby incorporated by reference in their entireties. Because of their usefulness in biotechnology, it is thus highly interesting to isolate new multimerization domains.

The multimerization activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including circular dichroism spectrum, gel filtration chromatography and thermal melting analyses.

In one embodiment, the invention relates to compositions and methods of using the protein of the invention or part thereof for preparing soluble multimeric proteins, which consist in multimers of fusion proteins containing a multimerization domain fused to a protein of interest, using any technique known to those skilled in the art including those described in international
5 patent WO9410308, which disclosure is hereby incorporated by reference in its entirety.

In another embodiment, the protein of the invention or part thereof or derivative thereof is used for detection and determination of an analyte in a biological liquid using the teachings of US patent 5,643,731, which disclosure is hereby incorporated by reference in its entirety. Briefly, a first multimerization domain is immobilized on a solid support and the second multimerization
10 domain is coupled to a specific binding partner for an analyte in a biological fluid. The two peptides are then brought into contact thereby immobilizing the binding partner on the solid phase. The biological sample is then contacted with the immobilized binding partner and the amount of analyte in the sample bound to the binding partner determined.

In still another embodiment, the protein of the invention or part thereof may be used to
15 construct multimerization devices comprising hybrid molecules with a functional domain fused to a multimerization domain in order to yield multimeric complexes with improved pharmacokinetic and pharmacological properties as described in WO0102440, which disclosure is hereby incorporated by reference in its entirety. In a preferred embodiment, the protein of the invention or part thereof may be used to construct different fusion proteins with different functional domains such as enzyme
20 moieties or cytotoxic moieties. Vectors encoding these different proteins may then be transfected in the same host cell in conditions allowing for multimerization, thus yielding multimeric multifunctional complexes.

Chaperone Protein of SEQ ID NO: 303 (internal designation D637548)

25 The protein of SEQ ID NO : 303 encoded by the cDNA of SEQ ID NO:134 is a chaperonin. Accordingly, the protein of SEQ ID NO:303 plays a role in protein synthesis/folding, cellular trafficking, and the cellular stress response. In addition, the protein of SEQ ID No: 303 has immunosuppressant and growth factor properties. It is able to depress delayed type hypersensitivity reactions. It is a product of primary and neoplastic cell proliferation and under these conditions acts
30 as a growth factor. It is also a product of platelet activation and may play a part in wound healing and skin repair. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:303 from positions 9 to 33, or from positions 7 to 101. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 303 having any of the biological activities described herein. The different activities of the protein of the invention or part thereof
35 may be assayed using any of the assays described in US 6,117,421 or any of the assays referred into US, 6117,421, which disclosures are hereby incorporated by reference in their entireties.

Chaperonins belong to a wider class of molecular chaperones, molecules involved in post-translational folding, targeting and assembly of other proteins, but which do not themselves form

part of the final assembled structure as discussed by Ellis et al., 1991, *Annu. Rev. Biochem.* 60 321-347, which disclosures are hereby incorporated by reference in their entireties. Most molecular chaperones are "heat shock" or "stress" proteins (hsp); i.e. their production is induced or increased by a variety of cellular insults (such as metabolic disruption, oxygen radicals, inflammation, infection and transformation), heat being only one of the better studies stresses as reviewed by Lindquist et al., 1988, *Annu. Rev. Genet.* 22 631-677, which disclosure is hereby incorporated by reference in its entirety. As well as these quantitative changes in specific protein levels, stress can induce the movement of constitutively produced stress proteins to different cellular compartments as referred to in the Lindquist reference mentioned above. The heat shock response is one of the most highly conserved genetic system known and the various heat shock protein families are among the most evolutionarily stable proteins in existence. The major stress proteins accumulate to very high levels in stressed cells but occur at low to moderate levels in cells that have not been stressed. As well as enabling cells to cope under adverse conditions, members of these families perform essential functions in normal cells.

Chaperones are also involved in a number of disorders, especially autoimmune diseases such as type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Sjogren syndrome, and mixed connective tissue disease (Feige et al. *EXS* 1996; 77:359-73; Feili-Hariri et al. *J Autoimmun* 2000; 14:133-42, which disclosures are hereby incorporated by reference in their entireties). Chaperones are also involved in various disorders including tuberculosis and leprosy (Zugel et al. *Clin Microbiol Rev* 1999; 12:19-39), neurogenerative disorders such as Alzheimer and Parkinson diseases (Yoo et al. *J Neural Transm Suppl* 1999; 57:315-22), and malignant disorders (Csermely et al. *Pharmacol Ther* 1998; 79:129-68), which disclosures are hereby incorporated by reference in their entireties.

In one embodiment, the protein of the invention or part thereof may be used to detect a potential pregnancy, preferably within 6-24 hours of fertilization using the teaching of Morton et al., 1976, *Proc. R. Soc. B.* 193 413-41 and US 6,117,421, which disclosures are hereby incorporated by reference in their entireties. Detection of the expression or activity of the protein of the invention may be performed using any techniques known to those skilled in the art including those described elsewhere in the application.

In another embodiment, molecules able to block the expression or activity of the protein of the invention, such as antibodies, antisense or triple helix oligonucleotides, dominant negative forms of the protein, polypeptides or small molecule inhibitors of the expression or activity of the proteins, may be used to induce abortion as described in US patent 6,117,421.

In still another embodiment, the protein of the invention or part thereof may be used to treat and/or prevent infertility and miscarriage using the simple administration of the protein of the invention or part thereof or using any of the gene therapy methods described elsewhere in the application and the teaching of the US patent 6,117,421.

In another embodiment, the present invention provide methods of using the present proteins to identify specific cell types in vitro and in vivo. For example, as chaperone proteins are often

upregulated in response to cellular stress, the detection of cells expressing elevated levels of the proteins provides a tool for detecting cells under stress. As cellular stress has been implicated in a number of disorders, such as cardiovascular disorders, neurodegenerative disorders, and cancer, the ability to detect such stress thus provides a diagnostic or screening tool for such conditions.

5 In addition, the present polypeptides and polynucleotides can be used to develop diagnostic and screening assays for diseases characterized by an abnormal level or activity of the protein of SEQ ID NO: 303 such as malignant disorders of various types, and autoimmune diseases including type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Sjogren syndrome, Graves disease, multiple sclerosis, and mixed connective tissue disease. Such assays can be performed
10 using any biological sample, such as serum or plasma.

In another embodiment, various disorders can be treated, attenuated and/or prevented by a protein of SEQ ID NO: 303, or part thereof, or any other compound that can affect the level or activity of the proteins such as nucleic acids, antibodies, or chemical substances. In a preferred embodiment, proteins or other compounds directed to the proteins of the invention can be used to
15 treat or prevent disorders in which the activity or level of the protein of SEQ ID NO: 303 is unbalanced. Such diseases include, but are not limited to, infectious diseases, neurodegenerative disorders as Alzheimer and Parkinson diseases, schizophrenia, alopecia, aging, atherosclerosis, malignant disorders of various types, and autoimmune diseases including type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, Sjogren syndrome, mixed connective tissue
20 disease, malignant disorders, autoimmune and any other neurodegenerative disorder. In another embodiment, the proteins of SEQ ID NO: 303 or part thereof can be used as vaccines for various disorders including, but not limited, to cancer (Wang et al. Immunol Invest 2000;29:131-7), tuberculosis (Silva et al. Microbes Infect 1999;1:429-35), diabetes (Int Immunol 1999;11:957-66), and atherosclerosis (Xu et al. Arterioscler Thromb 1992;12:789-99), which disclosures are hereby
25 incorporated by reference in their entireties.

One embodiment of the present invention relates to methods and compositions using the protein of SEQ ID NO:303 or fragments thereof as a stabilizing adjuvant to slow down protein degradation, boost the yields of recombinant proteins, prevent the aggregation of proteins or regenerate denatured proteins. In a preferred embodiment, the protein of SEQ ID NO:303 of
30 fragment thereof is mixed with a composition comprising the protein for which it is desired to slow down degradation, boost yield, or regenerate denatured proteins under conditions which facilitate the desired result. For example, numerous commercial assay kits commonly used by those skilled in the arts of molecular biology and biochemistry depend on the biological properties of proteins (mostly enzymes) which can be very short-lived in vitro due to the low stability of those proteins.
35 An example is described in Eur. Patent DE4124286, the disclosure of which is incorporated herein by reference in its entirety, wherein the low intrinsic stability of test solutions used in optical tests is increased by addition of chaperone proteins, thus making the test more sensitive. Another example is given in US patent 6,013,488, which disclosure is hereby incorporated by reference in its entirety,

wherein a heat-labile reverse transcriptase is able to perform cDNA synthesis at high temperature levels in the presence of a chaperone.

The protein of SEQ ID NO:303 may also be used to increase the yield or activity of recombinant proteins, preferably secreted proteins. In recombinant DNA technology, a major
5 unsolved problem is the solubility and biological activity of the recombinantly overexpressed protein in a host, especially a bacterial or yeast host. Many eukaryotic proteins, especially the secreted ones, require for correct folding a specific cellular machinery which is lacking in bacterial hosts such as *E. coli* or becomes insufficient in mammalian/yeast cells due to high expression of the protein. The ability of the protein of SEQ ID NO:303 or fragments thereof to ensure proper folding
10 of recombinant proteins may be utilized as follows. The protein of SEQ ID NO:303, or fragment thereof, may be coexpressed with the recombinant protein in bacterial or eukaryotic hosts to cause the hosts to express the heterologous proteins or polypeptides in a form having increased solubility and/or biological activity. For example, the protein of SEQ ID NO:303 or fragments thereof may be used in the methods described in US patent 5,773,245, the disclosure of which is incorporated
15 herein by reference in its entirety. Therefore the invention relates to a method for the correct folding, deaggregation or prevention of aggregation of a monomeric protein in vivo comprising: (a) constructing a host cell transformed with (i) a first DNA encoding a polypeptide having the amino acid sequence of a bioactive protein or a precursor thereof, wherein said polypeptide or precursor can aggregate within the cell to result in a multimeric, non-bioactive protein or precursor thereof
20 and (ii) a second DNA which enable the cell to co-express the protein of the invention or part thereof with the said polypeptide or precursor, (b) growing said host cell for sufficient time under conditions wherein said first DNA and said second DNA express said bioactive protein and said protein of the invention, respectively; and (c) obtaining monomeric protein that is a bioactive protein. Alternatively the protein of SEQ ID NO:303 or fragments thereof may be exogeneously
25 added to the cell cultures as described in PCT application WO 00/08135, the disclosure of which is incorporated herein by reference in its entirety.

The protein of SEQ ID NO:303 or fragments thereof may further be used to regenerate denatured proteins. Recombinantly expressed proteins with poor biological activity are routinely denatured with a potent denaturing agent, such as guanidine hydrochloride, followed by refolding
30 by dilution with a large amount of a diluent to reduce the concentration of the denaturing agent. However, this method often results in a poor refolding rate which may be significantly increased by addition of a cocktail of chaperone proteins in a fashion similar to that described in Eur. Patent EP0650975, the disclosure of which is incorporated herein by reference in its entirety. The advantage of using a cocktail of chaperone proteins is to accommodate differences in binding
35 specificity of the Hsp different families and the different members within each family.

In another embodiment of the present invention, the protein of SEQ ID NO:303 may be used to promote tissue repair and/or increase cell survival in stress conditions such as hypoxia, oxidative stress, genotoxic agents and more generally harmful conditions leading to programmed cell death. Those conditions include but are not limited to infarction, heart surgery, stroke,

neurodegenerative diseases, epilepsy, trauma, atherosclerosis, restenosis after angioplasty, and nerve damage.

In addition, the invention relates to compositions and methods for promoting cell growth both in vitro and in vivo using any of the techniques known to those skilled in the art including those described in the US patent 6,117,421. For example, soluble forms of the protein of the invention or part thereof may be added to cell culture medium in an amount effective to stimulate cell proliferation. Alternatively, any of the gene therapy methods described herein may be used to overexpress the protein of the invention or part thereof in vivo. Alternatively, the protein of the invention or part thereof may be directly administered in an amount effective to promoting cell growth in said subject. These applications are particularly important in individuals suffering from wounds or tissue damage to enhance tissue repair, in individuals to which organ or skin grafts have been applied, in individuals suffering from an inflammatory condition or an allergic disease.

In addition, the invention relates to compositions and methods for promoting immunosuppression in a subject using any of the techniques known to those skilled in the art including those described in the US patent 6,117,421. For example, the protein of the invention or part thereof may be directly administered in an amount effective to achieve immunosuppression in said subject. Alternatively, any of the gene therapy methods described herein may be used to overexpress the protein of the invention or part thereof in vivo. These applications are particularly important in cases in which immunosuppression is desired such as in individuals suffering from autoimmune disease including any of the diseases cited above and in individuals that have received an heterologous graft they could reject.

Ion transport Protein of SEQ ID NO: 276 (internal designation D538694)

The protein of SEQ ID NO : 276 encoded by the cDNA of SEQ ID NO:107 belongs to the FXDY family of small ion transport regulators or channels (Sweadner and Rael (2000) *Genomics* 68:41-56, which disclosure is hereby incorporated by reference in its entirety). The protein of SEQ ID NO: 276 or part thereof plays a role in the control of ion transport. Preferred polypeptides of the invention are polypeptides comprising the amino acids of SEQ ID NO:276 from positions 9 to 63, or from positions 16 to 29. Other preferred polypeptides of the invention are fragments of SEQ ID NO: 276 having any of the biological activities described herein. The activity of the protein of the invention or part thereof may be assayed using any of the assays known to those skilled in the art including those described in .

Osmoregulation occurs in all organisms, though the mechanisms differ according to the organism's environment. Fresh water inhabitants need to retain salts, whereas ocean inhabitants need to retain water. Terrestrial inhabitants need to conserve both water and salts. Organisms must balance these needs with a requirement to eliminate metabolic waste, such as nitrogenous waste, and generate secreted body fluids, such as saliva for digestion and sweat for thermoregulation.

In mammals, sweat glands, salivary glands, and the kidney all produce a primary secretion that is essentially isosmotic with blood and extracellular fluids. Modification of this primary secretion then occurs as much of the sodium chloride and water are reabsorbed as they pass through the excretory ducts of the glands and kidney, whereas potassium and bicarbonate ions are secreted.

- 5 This modification of the primary secretion is important in the sweat glands to conserve sodium chloride in hot environments, and in the salivary glands to conserve sodium chloride when excessive quantities of saliva are lost. This modification is critical in the kidney to maintain proper sodium and water balance in the extracellular fluids, a balance which also regulates arterial pressure. Loss of this modification activity by the duct cells causes a large loss of sodium and water,
10 resulting in severe dehydration and low blood volume, and ultimately to circulatory collapse.

- Sodium absorption by the intestines, especially in the colon, is necessary to prevent loss of sodium in the stools. The loss of sodium absorption produces a failure to absorb anions and water as well. The unabsorbed sodium chloride and water then lead to diarrhea, with further loss of sodium chloride from the body. Other body fluids may be under regulation similar to that seen in the
15 systems described above. For example, cerebrospinal fluid is produced by active sodium ion transport from the capillaries across the epithelium of the choroid plexus, which in turn attracts chloride ions and water. A counter flow of potassium and bicarbonate ions move out of the cerebrospinal fluid into the capillaries. A dysfunction in osmoregulation is associated with several disease states, including hyponatremia, renal failure, and hypernatremia. (Strange, K. (1992) J Am.
20 Soc. Nephrol. 3:12-27, which disclosure is hereby incorporated by reference in its entirety).

- In one embodiment, the protein of the invention may be useful in the diagnosis, prevention and/or treatment of osmoregulatory disorders including but not limited to diabetes insipidus, diarrhea, peritonitis, chronic renal failure, Addison's disease, syndrome of inappropriate antidiuretic hormone (SIADH), hypoaldosteronism, hyponatremia, adrenal insufficiency, hypothyroidism,
25 hypernatremia, hypokalemia, Barter's syndrome, Cushing's syndrome, metabolic acidosis, metabolic alkalosis, encephalopathy, edema, hypotension, and hypertension. For example, any of the gene therapy methods described herein may be used to overexpress the protein of the invention or part thereof in vivo. Alternatively, the protein of the invention or part thereof may be directly administered in an amount effective to promoting cell growth in said subject. For diagnostic
30 purposes, the expression of the protein of the invention could be investigated using any of the Northern blotting, RT-PCR or immunoblotting methods described herein and compared to the expression in control individuals. For prevention and/or treatment purposes, the protein of the invention may be used to enhance ion transport and prevent or treat osmoregulatory disorders using any of the gene therapy methods described herein.

35

Uses of antibodies

Antibodies of the present invention have uses that include, but are not limited to, methods known in the art to purify, detect, and target the polypeptides of the present invention including both *in vitro* and *in vivo* diagnostic and therapeutic methods. An example of such use using
5 immunoaffinity chromatography is given below. The antibodies of the present invention may be used either alone or in combination with other compositions. For example, the antibodies have use in immunoassays for qualitatively and quantitatively measuring levels of antigen-bearing substances, including the polypeptides of the present invention, in biological samples (*See, e.g.,* Harlow *et al.*, 1988). (Incorporated by reference in the entirety). The antibodies may also be used in therapeutic
10 compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

The invention further relates to antibodies that act as agonists or antagonists of the polypeptides of the present invention. For example, the present invention includes antibodies that disrupt the receptor/ligand interactions with the polypeptides of the invention either partially or fully. Included are both receptor-specific antibodies and ligand-specific antibodies. Included are
15 receptor-specific antibodies, which do not prevent ligand binding but prevent receptor activation. Receptor activation (i.e., signaling) may be determined by techniques described herein or otherwise known in the art. Also include are receptor-specific antibodies which both prevent ligand binding and receptor activation. Likewise, included are neutralizing antibodies that bind the ligand and prevent binding of the ligand to the receptor, as well as antibodies that bind the ligand, thereby
20 preventing receptor activation, but do not prevent the ligand from binding the receptor. Further included are antibodies that activate the receptor. These antibodies may act as agonists for either all or less than all of the biological activities affected by ligand-mediated receptor activation. The antibodies may be specified as agonists or antagonists for biological activities comprising specific activities disclosed herein. The above antibody agonists can be made using methods known in the
25 art. *See e.g.,* WO 96/40281; US Patent 5,811,097; Deng *et al.* (1998); Chen *et al.* (1998); Harrop *et al.* (1998); Zhu *et al.* (1998); Yoon *et al.* (1998); Prat *et al.* (1998); Pitard *et al.* (1997); Liautard *et al.* (1997); Carlson *et al.* (1997); Taryman *et al.* (1995); Muller *et al.* (1998); Bartunek *et al.* (1996) (said references incorporated by reference in their entireties).

As discussed above, antibodies of the polypeptides of the invention can, in turn, be utilized
30 to generate anti-idiotypic antibodies that “mimic” polypeptides of the invention using techniques well known to those skilled in the art (*See, e.g.* Greenspan and Bona (1989) and Nissinoff (1991), which disclosures are hereby incorporated by reference in their entireties). For example, antibodies which bind to and competitively inhibit polypeptide multimerization or binding of a polypeptide of the invention to ligand can be used to generate anti-idiotypes that “mimic” the polypeptide
35 multimerization or binding domain and, as a consequence, bind to and neutralize polypeptide or its ligand. Such neutralization anti-idiotypic antibodies can be used to bind a polypeptide of the invention or to bind its ligands/receptors, and thereby block its biological activity.

Immunoaffinity Chromatography

Antibodies prepared as described herein are coupled to a support. Preferably, the antibodies are monoclonal antibodies, but polyclonal antibodies may also be used. The support may be any of those typically employed in immunoaffinity chromatography, including Sepharose CL-4B

- 5 (Pharmacia, Piscataway, NJ), Sepharose CL-2B (Pharmacia, Piscataway, NJ), Affi-gel 10 (Biorad, Richmond, CA), or glass beads.

- The antibodies may be coupled to the support using any of the coupling reagents typically used in immunoaffinity chromatography, including cyanogen bromide. After coupling the antibody to the support, the support is contacted with a sample which contains a target polypeptide whose
- 10 isolation, purification or enrichment is desired. The target polypeptide may be a polypeptide selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool, variants and fragments thereof, or a fusion protein comprising said selected polypeptide or a fragment thereof.

- Preferably, the sample is placed in contact with the support for a sufficient amount of time
- 15 and under appropriate conditions to allow at least 50% of the target polypeptide to specifically bind to the antibody coupled to the support.

- Thereafter, the support is washed with an appropriate wash solution to remove polypeptides which have non-specifically adhered to the support. The wash solution may be any of those typically employed in immunoaffinity chromatography, including PBS, Tris-lithium chloride buffer
- 20 (0.1M lysine base and 0.5M lithium chloride, pH 8.0), Tris-hydrochloride buffer (0.05M Tris-hydrochloride, pH 8.0), or Tris/Triton/NaCl buffer (50mM Tris.cl, pH 8.0 or 9.0, 0.1% Triton X-100, and 0.5MNaCl).

- After washing, the specifically bound target polypeptide is eluted from the support using the high pH or low pH elution solutions typically employed in immunoaffinity chromatography. In
- 25 particular, the elution solutions may contain an eluant such as triethanolamine, diethylamine, calcium chloride, sodium thiocyanate, potassium bromide, acetic acid, or glycine. In some embodiments, the elution solution may also contain a detergent such as Triton X-100 or octyl-beta-D-glucoside.

EXPRESSION OF GENSET GENE PRODUCTS

30 **Spatial expression of the GENSET genes of the invention**

- Tissue expression of the cDNAs of the present invention was examined. Tables III and IV lists the number of hits for the cDNAs in Genset's libraries of tissues and cell types as well as in public databases. The tissues and cell types examined for polynucleotide expression were, for Table III: Brain; Fetal brain; Fetal kidney; Fetal liver; Pituitary gland; Liver; Placenta; Prostate;
- 35 Salivary gland; Stomach/Intestine; and Testis. For each cDNA referred to by its corresponding sequence identification number from the priority application (see Table I for corresponding SEQ ID NO in present application), the number of proprietary 5'ESTs (i.e. cDNA fragments) expressed in a

particular tissue referred to by its name is indicated in parentheses (second column). In addition, the bias in the spatial distribution of the polynucleotide sequences of the present invention was examined by comparing the relative proportions of the biological polynucleotides of a given tissue using the following statistical analysis. The under- or over-representation of a polynucleotide of a given cluster in a given tissue was performed using the normal approximation of the binomial distribution. When the observed proportion of a polynucleotide of a given tissue in a given consensus had less than 1% chance to occur randomly according to the chi² test, the frequency bias was reported as "preferred". The results are given in Table V as follows. For each polynucleotide showing a bias in tissue distribution as referred to by its sequence identification number in the first column, the list of tissues where the polynucleotides are under-represented is given in the second column entitled "low expression" and the list of tissues where the polynucleotides are over-represented is given in the third column entitled "high expression".

Evaluation of Expression Levels and Patterns of GENSET polypeptide-encoding mRNAs

The spatial and temporal expression patterns of GENSET polypeptide-encoding mRNAs, as well as their expression levels, may also be further determined as follows.

Expression levels and patterns of GENSET polypeptide-encoding mRNAs may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are hereby incorporated by reference. Briefly, a GENSET polynucleotide, or fragment thereof, corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the GENSET polynucleotide is at least a 100 nucleotides in length. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The GENSET polypeptide-encoding cDNAs, or fragments thereof, may also be tagged with nucleotide sequences for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A, the entire contents of which are incorporated by reference. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which it is desired to determine gene expression patterns. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an "anchoring enzyme," having a recognition site which is likely to be present at least once in most

cDNAs. The fragments which contain the 5' or 3' most region of the cleaved cDNA are isolated by binding to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a "tagging endonuclease" is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short "tag" fragments from the cDNAs. A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the "tagging endonuclease" to generate short "tag" fragments derived from the cDNAs in the second pool. The "tags" resulting from digestion of the first and second pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce "ditags." In some embodiments, the ditags are concatamerized to produce ligation products containing from 2 to 200 ditags. The tag sequences are then determined and compared to the sequences of the GENSET polypeptide-encoding cDNAs to determine which genes are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of a GENSET polypeptide-encoding gene in the cell, tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of GENSET gene expression may also be performed using arrays. For example, quantitative analysis of gene expression may be performed with GENSET polynucleotides, or fragments thereof in a complementary DNA microarray as described by Schena *et al.* (1995 and 1996) which disclosures are hereby incorporated by reference in their entireties. GENSET polypeptide-encoding cDNAs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C. Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm² microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1X SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1X SSC/0.2% SDS). Arrays are scanned in 0.1X SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with GENSET polypeptide-encoding cDNAs or fragments thereof in complementary DNA arrays as described by Pietu *et al.* (1996), which disclosure is hereby incorporated by reference in its entirety. The GENSET polynucleotides of the invention or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are

detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of GENSET genes can be done through high density nucleotide arrays as described by Lockhart *et al.* (1996) and Sosnowski *et al.* (1997), which disclosures are hereby incorporated by reference in their entireties. Oligonucleotides of 15-50 nucleotides corresponding to sequences of a GENSET polynucleotide or fragments thereof are synthesized directly on the chip (Lockhart *et al.*, supra) or synthesized and then addressed to the chip (Sosnowski *et al.*, supra). Preferably, the oligonucleotides are about 20 nucleotides in length. cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al.*, (supra) and application of different electric fields (Sosnowski *et al.*, supra), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of the GENSET polypeptide-encoding mRNA.

Uses of GENSET gene expression data

Once the expression levels and patterns of a GENSET polypeptide-encoding mRNA has been determined using any technique known to those skilled in the art, in particular those described in the section entitled "Evaluation of Expression Levels and Patterns of GENSET polypeptide-encoding mRNAs", or using the instant disclosure, these information may be used to design GENSET gene specific markers for detection, identification, screening and diagnosis purposes as well as to design DNA constructs with an expression pattern similar to a GENSET gene expression pattern.

25 Detection of GENSET polypeptide expression and/or biological activity

The invention further relates to methods of detection of GENSET polypeptide expression and/or biological activity in a biological sample using the polynucleotide and polypeptide sequences described herein. Such method can be used, for example, as a screen for normal or abnormal GENSET polypeptide expression and/or biological activity and, thus, can be used diagnostically. The biological sample for use in the methods of the present invention includes a suitable sample from, for example, a mammal, particularly a human. For example, the sample can be issued from tissues or cell lines having the same origin as tissues or cell lines in which the polypeptide is known to be expressed, e.g. using data from Tables III, IV, or V.

Detection of GENSET polypeptides

The invention further relates to methods of detection of GENSET polypeptide or encoding polynucleotides in a sample using the sequences described herein and any techniques known to

those skilled in the art. For example, a labeled polynucleotide probe having all or a functional portion of the nucleotide sequence of a GENSET polypeptide-encoding polynucleotide can be used in a method to detect a GENSET polypeptide-encoding polynucleotide in a sample. In one embodiment, the sample is treated to render the polynucleotides in the sample available for hybridization to a polynucleotide probe, which can be DNA or RNA. The resulting treated sample is combined with a labeled polynucleotide probe having all or a portion of the nucleotide sequence of the GENSET polypeptide-encoding cDNA or genomic sequence, under conditions appropriate for hybridization of complementary sequences to occur. Detection of hybridization of polynucleotides from the sample with the labeled nucleic probe indicates the presence of GENSET polypeptide-encoding polynucleotides in a sample. The presence of GENSET polypeptide-encoding mRNA is indicative of GENSET polypeptide-encoding gene expression.

Consequently, the invention comprises methods for detecting the presence of a polynucleotide comprising a nucleotide sequence selected from a group consisting of the sequences of SEQ ID NOs:1-169, 339-455, 561-784, the sequences of clone inserts of the deposited clone pool, sequences fully complementary thereto, fragments and variants thereof in a sample. In a first embodiment, said method comprises the following steps of:

- a) bringing into contact said sample and a nucleic acid probe or a plurality of nucleic acid probes which hybridize to said selected nucleotide sequence; and
- b) detecting the hybrid complex formed between said probe or said plurality of probes and said polynucleotide.

In a preferred embodiment of the above detection method, said nucleic acid probe or said plurality of nucleic acid probes is labeled with a detectable molecule. In another preferred embodiment of the above detection method, said nucleic acid probe or said plurality of nucleic acid probes has been immobilized on a substrate. In still another preferred embodiment, said nucleic acid probe or said plurality of nucleic acid probes has a sequence comprised in a sequence complementary to said selected sequence.

In a second embodiment, said method comprises the steps of:

- a) contacting said sample with amplification reaction reagents comprising a pair of amplification primers located on either side of the region of said nucleotide sequence to be amplified;
- b) performing an amplification reaction to synthesize amplification products containing said region of said selected nucleotide sequence; and
- c) detecting said amplification products.

In a preferred embodiment of the above detection method, when the polynucleotide to be amplified is a RNA molecule, preliminary reverse transcription and synthesis of a second cDNA strand are necessary to provide a DNA template to be amplified. In another preferred embodiment of the above detection method, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region. In still another preferred

embodiment, at least one of said amplification primer has a sequence comprised in said selected sequence or in the sequence complementary to said selected sequence.

Alternatively, a method of detecting GENSET polypeptide expression in a test sample can be accomplished using any product which binds to a GENSET olypeptide of the present invention
5 or a portion of a GENSET polypeptide. Such products may be antibodies, binding fragments of antibodies, polypeptides able to bind specifically to GENSET polypeptides or fragments thereof, including GENSET polypeptide agonists and antagonists. Detection of specific binding to the antibody indicates the presence of a GENSET polypeptide in the sample (e.g., ELISA).

Consequently, the invention is also directed to a method for detecting specifically the
10 presence of a GENSET polypeptide according to the invention in a biological sample, said method comprising the steps of:

- a) bringing into contact said biological sample with a product able to bind to a polypeptide of the invention or fragments thereof;
- b) allowing said product to bind to said polypeptide to form a complex; and
15 b) detecting said complex.

In a preferred embodiment of the above detection method, the product is an antibody. In a more preferred embodiment, said antibody is labeled with a detectable molecule. In another more preferred embodiment of the above detection method, said antibody has been immobilized on a substrate.

20 In addition, the invention also relates to methods of determining whether a GENSET gene product (e.g. a polynucleotide or polypeptide) is present or absent in a biological sample, said methods comprising the steps of:

- a) obtaining said biological sample from a human or non-human animal, preferably a mammal;
- 25 b) contacting said biological sample with a product able to bind to a GENSET polypeptide or encoding polynucleotide of the invention; and
- c) determining the presence or absence of said GENSET polypeptide-encoding gene product in said biological sample.

The present invention also relates to kits that can be used in the detection of GENSET
30 polypeptide-encoding gene expression products. The kit can comprise a compound that specifically binds a GENSET polypeptide (e.g. binding proteins, antibodies or binding fragments thereof (e.g. F(ab')₂ fragments) or a GENSET polypeptide-encoding mRNA (e.g. a complementary probe or primer), for example, disposed within a container means. The kit can further comprise ancillary reagents, including buffers and the like.

35 *Detection of GENSET polypeptide biological activity*

The invention further includes methods of detecting specifically a GENSET polypeptide biological activity, and to identify compounds capable of modulating the activity of a GENSET polypeptide. Assessing the GENSET polypeptide biological activity may be performed by the

detection of a change in any cellular property associated with the GENSET polypeptide, using a variety of techniques, including those described herein. To identify modulators of the polypeptides, a control is preferably used. For example, a control sample includes all of the same reagents but lacks the compound or agent being assessed; it is treated in the same manner as the test sample. A
5 number of potentially assayable biological activities for many of the herein-described proteins are described *supra*, under the heading, "Uses of polypeptides of the invention."

The present invention also relates to kits that can be used in the detection of GENSET polypeptide biological activity. The kit can comprise, e.g. substrates for GENSET polypeptides, GENSET-binding compounds, antibodies to GENSET polypeptides, etc., for example, disposed
10 within a container means. The kit can further comprise ancillary reagents, including buffers and the like.

Identification of a specific context of GENSET polypeptide-encoding gene expression

When the expression pattern of a GENSET polypeptide-encoding mRNA shows that a GENSET polypeptide-encoding gene is specifically expressed in a given context, probes and
15 primers specific for this gene as well as antibodies binding to the GENSET polypeptide-encoding polynucleotide may then be used as markers for the specific context. Examples of specific contexts are: specific expression in a given tissue/cell or tissue/cell type (see, e.g., Tables III-V), expression at a given stage of development of a process such as embryo development or disease development, or specific expression in a given organelle. Such primers, probes, and antibodies are useful
20 commercially to identify tissues/cells/organelles of unknown origin, for example, forensic samples, differentiated tumor tissue that has metastasized to foreign bodily sites, or to differentiate different tissue types in a tissue cross-section using any technique known to those skilled in the art including *in situ* PCR or immunochemistry for example.

For example, the cDNAs and proteins of the sequence listing and fragments thereof, may be
25 used to distinguish human tissues/cells from non-human tissues/cells and to distinguish between human tissues/cells/organelles that do and do not express the polynucleotides comprising the cDNAs. By knowing the expression pattern of a given GENSET polypeptide, either through routine experimentation or by using the instant disclosure, the polynucleotides and polypeptides of the present invention may be used in methods of determining the identity of an unknown tissue/cell
30 sample/organelle. As part of determining the identity of an unknown tissue/cell sample/organelle, the polynucleotides and polypeptides of the present invention may be used to determine what the unknown tissue/cell sample is and what the unknown sample is not. For example, if a cDNA is expressed in a particular tissue/cell type/organelle, and the unknown tissue/cell sample/organelle does not express the cDNA, it may be inferred that the unknown tissue/cells are either not human or
35 not the same human tissue/cell type/organelle as that which expresses the cDNA. These methods of determining tissue/cell/organelle identity are based on methods which detect the presence or absence of the mRNA (or corresponding cDNA) in a tissue/cell sample using methods well known in the art (e.g., hybridization, PCR based methods, immunoassays, immunochemistry, ELISA).

Examples of such techniques are described in more detail below. Therefore, the invention encompasses uses of the polynucleotides and polypeptides of the invention as tissue markers. In a preferred embodiment, polynucleotides preferentially expressed in given tissues as indicated in Tables III-V and polypeptides encoded by such polynucleotides are used for this purpose. The invention also encompasses uses of polypeptides of the invention as organelle markers.

Consequently, the present invention encompasses methods of identification of a tissue/cell type/subcellular compartment, wherein said method includes the steps of:

- a) contacting a biological sample which identity is to be assayed with a product able to bind a GENSET gene product; and
- 10 b) determining whether a GENSET gene product is expressed in said biological sample.

Products that are able to bind specifically to a GENSET gene product, namely a GENSET polypeptide or a GENSET polypeptide-encoding mRNA, include GENSET polypeptide binding proteins, antibodies or binding fragments thereof (e.g. F(ab')₂ fragments), as well as GENSET polynucleotide complementary probes and primers.

Step b) may be performed using any detection method known to those skilled in the art including those disclosed herein, especially in the section entitled "Detection of GENSET polypeptide expression and/or biological activity".

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

20 Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations which are conjugated, directly (e.g., green fluorescent protein) or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

25 Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble
30 immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, (1980) or Rose *et al.*, (1980), which disclosures
35 are hereby incorporated by reference in their entirety.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as

horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific anti-tissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I , and detected by overlaying the antibody treated preparation with photographic emulsion. Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required. Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm , unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer. Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed. If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available. The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of Tissue Specific Soluble Proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection. A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis. A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis *et al.*, Section 19-2 (1986), using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5 to 55

ul, and containing from about 1 to 100 ug protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis *et al.*, (1986) Section 19-3. One set of nitrocellulose blots is stained with

- 5 Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described herein. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

- In either procedure A or B, a detectable label can be attached to the primary tissue antigen-
 10 primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive
 15 protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody. The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

20 Screening and diagnosis of abnormal GENSET polypeptide expression and/or biological activity

- Moreover, antibodies and/or primers specific for GENSET polypeptide expression may also be used to identify abnormal GENSET polypeptide expression and/or biological activity, and subsequently to screen and/or diagnose disorders associated with abnormal GENSET polypeptide expression. For example, a particular disease may result from lack of expression, over expression,
 25 or under expression of a GENSET polypeptide-encoding mRNA. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disorder, genes responsible for this disorder may be identified. Primers, probes and antibodies specific for this GENSET polypeptide may then be used to elaborate kits of screening and diagnosis for a disorder in which the gene of interest is specifically expressed or in
 30 which its expression is specifically dysregulated, i.e. underexpressed or overexpressed.

Screening for specific disorders

- The present invention also relates to methods and uses of GENSET polypeptides for identifying individuals having elevated or reduced levels of GENSET polypeptides, which individuals are likely to benefit from therapies to suppress or enhance GENSET polypeptide-
 35 encoding gene expression, respectively. One example of such methods and uses comprises the steps of:

- a) obtaining from a mammal a biological sample;

b) detecting the presence in said sample of a GENSET polypeptide-encoding gene product (mRNA or protein);

c) comparing the amount of said GENSET polypeptide-encoding gene product present in said sample with that of a control sample; and

5 d) determining whether said human or non-human mammal has a reduced or elevated level of GENSET gene expression compared to the control sample.

A biological sample from a subject affected by, or at risk of developing, any disease or condition associated with a GENSET polypeptide can be screened for the presence of increased or decreased levels of GENSET gene product, relative to a normal population (standard or control),
10 with an increased or decreased level of the GENSET polypeptide relative to the normal population being indicative of predisposition to or a present indication of the disease or condition, or any symptom associated with the disease or condition. Such individuals would be candidates for therapies, e.g., treatment with pharmaceutical compositions comprising the GENSET polypeptide, a polynucleotide encoding the GENSET polypeptide, or any other compound that affects the
15 expression or activity of the GENSET polypeptide. Generally, the identification of elevated levels of the GENSET polypeptide in a patient would be indicative of an individual that would benefit from treatment with agents that suppress GENSET polypeptide expression or activity, and the identification of low levels of the GENSET polypeptide in a patient would be indicative of an individual that would benefit from agents that induce GENSET expression or activity.

20 Biological samples suitable for use in this method include any biological fluids, including, but not limited to, blood, saliva, milk, and urine. Tissue samples (e.g. biopsies) can also be used in the method of the invention, including samples derived from any tissue associated with GENSET gene expression (see, e.g. Tables III-V). Cell cultures or cell extracts derived, for example, from tissue biopsies can also be used. The detection step of the present method can be performed using
25 standard protocols for protein/mRNA detection. Examples of suitable protocols include Northern blot analysis, immunoassays (e.g. RIA, Western blots, immunohistochemical analyses), and PCR.

Thus, the present invention further relates to methods and uses of GENSET polypeptides for identifying individuals or non-human animals at increased risk for developing, or present state of having, certain diseases/disorders associated with abnormal GENSET polypeptide expression or
30 biological activity. One example of such methods comprises the steps of:

a) obtaining from a human or non-human mammal a biological sample;

b) detecting the presence in said sample of a GENSET gene product (mRNA or protein);

c) comparing the amount of said GENSET gene product present in said sample with that of a control sample; and

35 d) determining whether said human or non-human mammal is at increased risk for developing, or present state of having, a diseases or disorder.

In preferred embodiments, the biological sample is taken from animals presenting any symptom associated with any disease or condition associated with a GENSET gene product. In accordance with this method, the presence in the sample of altered (e.g. increased or decreased)

levels of the GENSET product indicates that the subject is predisposed to the disease or condition. Biological samples suitable for use in this method include biological fluids including, but not limited to, blood, saliva, milk, and urine. Tissue samples (e.g. biopsies) can also be used in the method of the invention, including samples derived from any of the tissues listed in Tables III-V.

- 5 Cell cultures or cell extracts derived, for example, from tissue biopsies can also be used.

The diagnostic methodologies described herein are applicable to both humans and non-human mammals.

Detection of GENSET gene mutations

- The invention also encompasses methods and uses of GENSET polynucleotides to detect mutations in GENSET polynucleotides of the invention. Such methods may advantageously be used to detect mutations occurring in GENSET genes and preferably in their regulatory regions. When the mutation was proven to be associated with a disease, the detection of such mutations may be used for screening and diagnosis purposes.

- In one embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in GENSET genes and preferably in their regulatory regions. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant, mutations on the GENSET genes that have been identified according, for example to the technique used by Huang *et al.*(1996) or Samson *et al.*(1996), which disclosures are hereby incorporated by reference in their entireties.

- Another technique that is used to detect mutations in GENSET genes is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of a GENSET genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference wild gene sequence of the GENSET gene. In one such design, termed 4L tiled array, is implemented a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes flanking a mutation position. This technique was described by Chee *et al.* in 1996, which disclosure is hereby incorporated by reference in its entirety.

Construction of DNA constructs with a GENSET gene expression pattern

In addition, characterization of the spatial and temporal expression patterns and expression levels of GENSET polypeptide-encoding mRNAs is also useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as
5 discussed below.

DNA Constructs That Direct Temporal And Spatial GENSET Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.

In order to study the physiological and phenotypic consequences of a lack of synthesis of a GENSET polypeptide, both at the cellular level and at the multi cellular organism level, the
10 invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of a GENSET polypeptide-encoding genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to a nucleotide sequence selected from the group consisting of sequences of SEQ ID NOs:1-169, 339-455, 561-784 and sequences of clone inserts of the deposited
15 clone pool, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in the 5'-regulatory sequence or in an exon of the GENSET polypeptide-encoding genomic sequence or within the GENSET polypeptide-encoding cDNA.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E.*
20 *coli* transposon Tn10 for controlling the GENSET gene expression, such as described by Gossen *et al.* (1992, 1995) and Furth *et al.* (1994), which disclosures are hereby incorporated by reference in their entireties. Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to either a minimal promoter or a 5'-regulatory sequence of the GENSET gene, said minimal promoter or said GENSET polynucleotide regulatory sequence being operably linked to a
25 polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a GENSET polypeptide, or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed
30 under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention comprise both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor. In a specific embodiment, the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of
35 interest is silent in the absence of tetracycline and induced in its presence.

DNA Constructs Allowing Homologous Recombination: Replacement Vectors

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is found in the GENSET polypeptide-encoding genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycin resistance (*neo*); and (c) a second nucleotide sequence that is found in the GENSET polypeptide-
5 encoding genomic sequence, and is located on the genome downstream the first GENSET polypeptide-encoding nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream of the nucleotide sequence (a) or downstream from the nucleotide sequence (c). Preferably, the negative selection marker comprises the thymidine kinase (*tk*) gene (Thomas *et al.*,
10 1986), the hygromycin beta gene (Te Riele *et al.*, 1990), the *hprt* gene (Van der Lugt *et al.*, 1991; Reid *et al.*, 1990) or the Diphtheria toxin A fragment (*Dt-A*) gene (Nada *et al.*, 1993; Yagi *et al.* 1990), which disclosures are hereby incorporated by reference in their entireties. Preferably, the positive selection marker is located within a GENSET exon sequence so as to interrupt the sequence encoding a GENSET polypeptide. These replacement vectors are described, for example, by
15 Thomas *et al.* (1986; 1987), Mansour *et al.* (1988) and Koller *et al.* (1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a GENSET polypeptide-encoding regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more
20 preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *loxP* site. The *loxP* site is composed of two palindromic sequences of 13 bp separated by
25 a 8 bp conserved sequence (Hoess *et al.*, 1986), which disclosure is hereby incorporated by reference in its entirety. The recombination by the Cre enzyme between two *loxP* sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*loxP* system used in combination with a homologous recombination technique has been first described by Gu *et al.* (1993, 1994), which disclosures are hereby incorporated by
30 reference in their entireties. Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *loxP* sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the
35 recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki *et al.* (1995), which disclosure is hereby incorporated by reference in its entirety, or by lipofection of the enzyme into the cells, such as described by Baubonis *et al.* (1993), which disclosure is hereby incorporated by reference in its

entirety; (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu *et al.* (1993) and Sauer *et al.* (1988), which disclosures are hereby incorporated by reference in their
5 entirety; (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu *et al.* (1994).

10 In a specific embodiment, the vector containing the sequence to be inserted in the GENSET gene by homologous recombination is constructed in such a way that selectable markers are flanked by *loxP* sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the GENSET sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection
15 marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*loxP* system are described by Zou *et al.* (1994), which disclosure is hereby incorporated by reference in its entirety.

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end:
(a) a first nucleotide sequence that is comprised in the GENSET genomic sequence; (b) a nucleotide
20 sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *loxP* site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the GENSET genomic sequence, and is located on the genome downstream of the first GENSET nucleotide sequence (a).

25 The sequences defining a site recognized by a recombinase, such as a *loxP* site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *loxP* sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

30 In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two *loxP* sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence
35 and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu *et al.* (1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result from the breeding of two transgenic animals, the first transgenic animal bearing the GENSET-derived sequence of interest containing the *loxP* sites as described above and the second transgenic

animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu *et al.*(1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae *et al.* (1995), which disclosures are hereby incorporated by reference in their entireties.

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a GENSET genomic sequence or a GENSET cDNA sequence, and most preferably an altered copy of a GENSET genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination).

15 **MODIFYING GENSET POLYOPTIDE EXPRESSION AND/OR BIOLOGICAL ACTIVITY**

Modifying endogenous GENSET expression and/or biological activity is expressly contemplated by the present invention.

Screening for compounds that modulate GENSET expression and/or biological activity

The present invention further relates to compounds able to modulate GENSET expression and/or biological activity and methods to use these compounds. Such compounds may interact with the regulatory sequences of GENSET genes or they may interact with GENSET polypeptides directly or indirectly.

Compounds Interacting With GENSET Regulatory Sequences

The present invention also concerns a method for screening substances or molecules that are able to interact with the regulatory sequences of a GENSET gene, such as for example promoter or enhancer sequences in untranscribed regions of the genomic DNA, as determined using any techniques known to those skilled in the art including those described in the section entitled "Identification of Promoters in Cloned Upstream Sequences, or such as regulatory sequences located in untranslated regions of GENSET mRNA.

Sequences within untranscribed or untranslated regions of polynucleotides of the invention may be identified by comparison to databases containing known regulatory sequence such as transcription start sites, transcription factor binding sites, promoter sequences, enhancer sequences, 5'UTR and 3'UTR elements (Pesole *et al.*, 2000; <http://igs-server.cnrs-mrs.fr/~gauthere/UTR/index.html>). Alternatively, the regulatory sequences of interest may be identified through conventional mutagenesis or deletion analyses of reporter plasmids using, for

instance, techniques described in the section entitled "Identification of Promoters in Cloned Upstream Sequences".

Following the identification of potential GENSET regulatory sequences, proteins which interact with these regulatory sequences may be identified as described below.

- 5 Gel retardation assays may be performed independently in order to screen candidate molecules that are able to interact with the regulatory sequences of the GENSET gene, such as described by Fried and Crothers (1981), Garner and Revzin (1981) and Dent and Latchman (1993), the teachings of these publications being herein incorporated by reference. These techniques are based on the principle according to which a DNA or mRNA fragment which is bound to a protein
10 migrates slower than the same unbound DNA or mRNA fragment. Briefly, the target nucleotide sequence is labeled. Then the labeled target nucleotide sequence is brought into contact with either a total nuclear extract from cells containing regulation factors, or with different candidate molecules to be tested. The interaction between the target regulatory sequence of the GENSET gene and the candidate molecule or the regulation factor is detected after gel or capillary electrophoresis through
15 a retardation in the migration.

- Nucleic acids encoding proteins which are able to interact with the promoter sequence of the GENSET gene, more particularly a nucleotide sequence selected from the group consisting of the polynucleotides of the 5' and 3' regulatory region or a fragment or variant thereof, may be identified by using a one-hybrid system, such as that described in the booklet enclosed in the
20 Matchmaker One-Hybrid System kit from Clontech (Catalog Ref. no. K1603-1, the technical teachings of which are herein incorporated by reference). Briefly, the target nucleotide sequence is cloned upstream of a selectable reporter sequence and the resulting polynucleotide construct is integrated in the yeast genome (*Saccharomyces cerevisiae*). Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem. The yeast cells containing the
25 reporter sequence in their genome are then transformed with a library comprising fusion molecules between cDNAs encoding candidate proteins for binding onto the regulatory sequences of the GENSET gene and sequences encoding the activator domain of a yeast transcription factor such as GAL4. The recombinant yeast cells are plated in a culture broth for selecting cells expressing the reporter sequence. The recombinant yeast cells thus selected contain a fusion protein that is able to
30 bind onto the target regulatory sequence of the GENSET gene. Then, the cDNAs encoding the fusion proteins are sequenced and may be cloned into expression or transcription vectors *in vitro*. The binding of the encoded polypeptides to the target regulatory sequences of the GENSET gene may be confirmed by techniques familiar to the one skilled in the art, such as gel retardation assays or DNase protection assays.

35 Ligands interacting with GENSET polypeptides

 For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to a GENSET protein

or one of its fragments or variants or to modulate the expression of the polynucleotide coding for GENSET or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of a GENSET protein is brought into contact with the corresponding purified GENSET protein, for example the corresponding purified recombinant GENSET protein produced by a recombinant cell host as described herein, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool, with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang *et al.* (1997) or the affinity capillary electrophoresis method described by Bush *et al.* (1997), the disclosures of which are incorporated by reference, can be used.

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with a GENSET protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent, radioactive, or enzymatic tag and placed in contact with immobilized GENSET protein, or a fragment thereof under conditions which permit specific binding to occur. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

Various candidate substances or molecules can be assayed for interaction with a GENSET polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule comprises a polypeptide, this polypeptide may be the resulting expression product of a phage clone belonging to a phage-based random peptide library, or alternatively the polypeptide may be the resulting expression product of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

A. Candidate ligands obtained from random peptide libraries

In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode for peptides of 8 to 20 amino acids in length (Oldenburg *et al.*, 1992; Valadon *et al.*, 1996; Lucas, 1994; Westerink, 1995; Felici *et al.*, 1991), which disclosures are hereby incorporated by reference in their entireties.

According to this particular embodiment, the recombinant phages expressing a protein that binds to an immobilized GENSET protein is retained and the complex formed between the GENSET protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the GENSET protein.

5 Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized GENSET protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the GENSET protein are then eluted by a buffer (acid pH) or immunoprecipitated by the monoclonal antibody produced by the hybridoma anti-GENSET, and this phage population is
10 subsequently amplified by an over-infection of bacteria (for example *E. coli*). The selection step may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step comprises characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the phage insert in another host-vector system, or sequencing the insert contained in the
15 selected recombinant phages.

B. Candidate ligands obtained by competition experiments.

Alternatively, peptides, drugs or small molecules which bind to a GENSET protein or fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide
20 selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool, may be identified in competition experiments. In such assays, the GENSET protein, or a fragment thereof, is immobilized to a surface, such as a plastic plate. Increasing amounts of the peptides, drugs or small molecules are placed in contact with the immobilized GENSET protein, or a fragment thereof, in
25 the presence of a detectable labeled known GENSET protein ligand. For example, the GENSET ligand may be detectably labeled with a fluorescent, radioactive, or enzymatic tag. The ability of the test molecule to bind the GENSET protein, or a fragment thereof, is determined by measuring the amount of detectably labeled known ligand bound in the presence of the test molecule. A decrease in the amount of known ligand bound to the GENSET protein, or a fragment thereof, when
30 the test molecule is present indicated that the test molecule is able to bind to the GENSET protein, or a fragment thereof.

C. Candidate ligands obtained by affinity chromatography.

Proteins or other molecules interacting with a GENSET protein, or a fragment thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more
35 preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool, can also be found using affinity columns

- which contain the GENSET protein, or a fragment thereof. The GENSET protein, or a fragment thereof, may be attached to the column using conventional techniques including chemical coupling to a suitable column matrix such as agarose, Affi Gel®, or other matrices familiar to those of skill in art. In some embodiments of this method, the affinity column contains chimeric proteins in
- 5 which the GENSET protein, or a fragment thereof, is fused to glutathion S transferase (GST). A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins or other molecules interacting with the GENSET protein, or a fragment thereof, attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al.* (1997), the disclosure of which is incorporated by reference.
- 10 Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

D. Candidate ligands obtained by optical biosensor methods

- Proteins interacting with a GENSET protein, or a fragment comprising a contiguous span of
- 15 at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool, can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo *et al.* (1995), the disclosures of which are incorporated
- 20 by reference. This technique permits the detection of interactions between molecules in real time, without the need of labeled molecules. This technique is based on the surface plasmon resonance (SPR) phenomenon. Briefly, the candidate ligand molecule to be tested is attached to a surface (such as a carboxymethyl dextran matrix). A light beam is directed towards the side of the surface that does not contain the sample to be tested and is reflected by said surface. The SPR phenomenon
- 25 causes a decrease in the intensity of the reflected light with a specific association of angle and wavelength. The binding of candidate ligand molecules cause a change in the refraction index on the surface, which change is detected as a change in the SPR signal. For screening of candidate ligand molecules or substances that are able to interact with the GENSET protein, or a fragment thereof, the GENSET protein, or a fragment thereof, is immobilized onto a surface. This surface
- 30 comprises one side of a cell through which flows the candidate molecule to be assayed. The binding of the candidate molecule on the GENSET protein, or a fragment thereof, is detected as a change of the SPR signal. The candidate molecules tested may be proteins, peptides, carbohydrates, lipids, or small molecules generated by combinatorial chemistry. This technique may also be performed by immobilizing eukaryotic or prokaryotic cells or lipid vesicles exhibiting an
- 35 endogenous or a recombinantly expressed GENSET protein at their surface.

The main advantage of the method is that it allows the determination of the association rate between the GENSET protein and molecules interacting with the GENSET protein. It is thus

possible to select specifically ligand molecules interacting with the GENSET protein, or a fragment thereof, through strong or conversely weak association constants.

E. Candidate ligands obtained through a two-hybrid screening assay.

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields
5 and Song, 1989), which disclosure is hereby incorporated by reference in its entirety, and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in the US Patent No. US 5,667,973 and the US Patent No. 5,283,173, the technical teachings of both patents being herein incorporated by reference.

The general procedure of library screening by the two-hybrid assay may be performed as
10 described by Harper *et al.* (1993) or as described by Cho *et al.* (1998) or also Fromont-Racine *et al.* (1997), which disclosures are hereby incorporated by reference in their entireties.

The bait protein or polypeptide comprises, consists essentially of, or consists of a GENSET polypeptide or a *fragment* thereof comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids
15 of a polypeptide selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool.

More precisely, the nucleotide sequence encoding the GENSET polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example
20 pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "prey" polypeptides.

25 A third vector contains a detectable marker gene, such as beta galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the
30 two different yeast strains may be the followings :

- Y190, the phenotype of which is (MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS *GAL*-HIS3, *cyh*^r);

- Y187, the phenotype of which is (MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmet^r), which is the opposite mating type of Y190.

35 Briefly, 20 µg of pAS2/GENSET and 20 µg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine synthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive

colonies (His⁺, beta-gal⁺) are then grown on plates lacking histidine, leucine, but containing tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/GENSET plasmids but retention of pACT-cDNA library plasmids. The resulting Y190 strains are mated with Y187 strains expressing GENSET or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as Gal4 fusions as described by Harper *et al.* (1993) and by Bram *et al.* (1993), which disclosures are hereby incorporated by reference in their entireties, and screened for beta galactosidase by filter lift assay. Yeast clones that are beta gal- after mating with the control Gal4 fusions are considered false positives.

In another embodiment of the two-hybrid method according to the invention, interaction between the GENSET or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the kit, the disclosure of which is incorporated herein by reference, nucleic acids encoding the GENSET protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between GENSET and the protein or peptide encoded by the initially selected cDNA insert.

Compounds Modulating GENSET biological activity

Another method of screening for compounds that modulate GENSET expression and/or biological activity is by measuring the effects of test compounds on specific biological activity, e.g. a GENSET biological activity in a host cell. In one embodiment, the present invention relates to a method of identifying an agent which alters GENSET biological activity, wherein a nucleic acid construct comprising a nucleic acid which encodes a mammalian GENSET polypeptide is introduced into a host cell. The host cells produced are maintained under conditions appropriate for expression of the encoded mammalian GENSET polypeptides, whereby the nucleic acid is expressed. The host cells are then contacted with a compound to be assessed (an "agent," or "test agent"), and the properties of the cells is assessed. Detection of a change in any GENSET polypeptide-associated property in the presence of the agent indicates that the agent alters GENSET activity. In a particular embodiment, the invention relates to a method of identifying an agent which is an activator of GENSET activity, wherein detection of an increase of any GENSET polypeptide-associated property in the presence of the agent indicates that the agent activates GENSET activity. In another particular embodiment, the invention relates to a method of identifying an agent which is an inhibitor of GENSET activity, wherein detection of a decrease of

any GENSET polypeptide-associated property in the presence of the agent indicates that the agent inhibits GENSET activity.

In a particular embodiment, a high throughput screen can be used to identify agents that activate (enhance) or inhibit GENSET activity (See e.g., PCT publication WO 98/45438, which disclosure is hereby incorporated by reference in its entirety). For example, the method of identifying an agent which alters GENSET activity can be performed as follows. A nucleic acid construct comprising a polynucleotide which encodes a mammalian GENSET polypeptide is introduced into a host cell to produce recombinant host cells. The recombinant host cells are then maintained under conditions appropriate for expression of the encoded mammalian GENSET polypeptide, whereby the nucleic acid is expressed. The compound to be assessed is added to the recombinant host cells; the resulting combination is referred to as a test sample. A detectable, GENSET polypeptide-associated property of the cells is detected. A control can be used in the methods of detecting agents which alter GENSET activity. For example, the control sample includes the same reagents but lacks the compound or agent being assessed; it is treated in the same manner as the test sample.

Methods of Screening for Compounds Modulating GENSET Expression and/or Activity

The present invention also relates to methods of screening compounds for their ability to modulate (e.g. increase or inhibit) the activity or expression of GENSET. More specifically, the present invention relates to methods of testing compounds for their ability either to increase or to decrease expression or activity of GENSET. The assays are performed *in vitro* or *in vivo*.

In vitro methods

In vitro, cells expressing GENSET polypeptides are incubated in the presence and absence of the test compound. By determining the level of GENSET expression in the presence of the test compound or the level of GENSET activity in the presence of the test compound, compounds can be identified that suppress or enhance GENSET expression or activity. Alternatively, constructs comprising a GENSET regulatory sequence operably linked to a reporter gene (e.g. luciferase, chloramphenicol acetyl transferase, LacZ, green fluorescent protein, etc.) can be introduced into host cells and the effect of the test compounds on expression of the reporter gene detected. Cells suitable for use in the foregoing assays include, but are not limited to, cells having the same origin as tissues or cell lines in which the polypeptide is known to be expressed using the data from Tables III, IV, or V.

Consequently, the present invention encompasses a method for screening molecules that modulate the expression of a GENSET gene, said screening method comprising the steps of:

- a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding a GENSET protein or a variant or a fragment thereof, placed under the control of its own promoter;
- b) bringing into contact said cultivated cell with a molecule to be tested;

c) quantifying the expression of said GENSET protein or a variant or a fragment thereof in the presence of said molecule.

Using DNA recombination techniques well known by the one skilled in the art, the GENSET protein encoding DNA sequence is inserted into an expression vector, downstream from its promoter sequence. As an illustrative example, the promoter sequence of the GENSET gene is contained in the 5' untranscribed region of the GENSET genomic DNA.

The quantification of the expression of a GENSET protein may be realized either at the mRNA level (using for example Northern blots, RT-PCR, preferably quantitative RT-PCR with primers and probes specific for the GENSET mRNA of interest) or at the protein level (using polyclonal or monoclonal antibodies in immunoassays such as ELISA or RIA assays, Western blots, or immunochemistry).

The present invention also concerns a method for screening substances or molecules that are able to increase, or in contrast to decrease, the level of expression of a GENSET gene. Such a method may allow the one skilled in the art to select substances exerting a regulating effect on the expression level of a GENSET gene and which may be useful as active ingredients included in pharmaceutical compositions for treating patients suffering from disorders associated with abnormal levels of GENSET products.

Thus, another part of the present invention is a method for screening a candidate molecule that modulates the expression of a GENSET gene, this method comprises the following steps:

- a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises a GENSET 5' regulatory region or a regulatory active fragment or variant thereof, operably linked to a polynucleotide encoding a detectable protein;
- b) obtaining a candidate molecule; and
- c) determining the ability of said candidate molecule to modulate the expression levels of said polynucleotide encoding the detectable protein.

In a further embodiment, said nucleic acid comprising a GENSET 5' regulatory region or a regulatory active fragment or variant thereof, includes the 5'UTR region of a GENSET cDNA selected from the group comprising of the 5'UTRs of the sequences of SEQ ID NOs:1-169, 339-455, 561-784, sequences of clones inserts of the deposited clone pool, regulatory active fragments and variants thereof. In a more preferred embodiment of the above screening method, said nucleic acid includes a promoter sequence which is endogenous with respect to the GENSET 5'UTR sequence. In another more preferred embodiment of the above screening method, said nucleic acid includes a promoter sequence which is exogenous with respect to the GENSET 5'UTR sequence defined therein.

Preferred polynucleotides encoding a detectable protein are polynucleotides encoding beta galactosidase, green fluorescent protein (GFP) and chloramphenicol acetyl transferase (CAT).

The invention further relates to a method for the production of a pharmaceutical composition comprising a method of screening a candidate molecule that modulates the expression

of a GENSET gene and furthermore mixing the identified molecule with a pharmaceutically acceptable carrier.

The invention also pertains to kits for the screening of a candidate substance modulating the expression of a GENSET gene. Preferably, such kits comprise a recombinant vector that allows the expression of a GENSET 5' regulatory region or a regulatory active fragment or a variant thereof, operably linked to a polynucleotide encoding a detectable protein or a GENSET protein or a fragment or a variant thereof. More preferably, such kits include a recombinant vector that comprises a nucleic acid including the 5'UTR region of a GENSET cDNA selected from the group comprising the 5'UTRs of the sequences of SEQ ID NOs: 1-169, 339-455, 561-784, sequences of clone inserts of the deposited clone pool, regulatory active fragments and variants thereof, being operably linked to a polynucleotide encoding a detectable protein.

For the design of suitable recombinant vectors useful for performing the screening methods described above, it will be referred to the section of the present specification wherein the preferred recombinant vectors of the invention are detailed.

Another object of the present invention comprises methods and kits for the screening of candidate substances that interact with a GENSET polypeptide, fragments or variants thereof. By their capacity to bind covalently or non-covalently to a GENSET protein, fragments or variants thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*.

In vitro, said interacting molecules may be used as detection means in order to identify the presence of a GENSET protein in a sample, preferably a biological sample.

A method for the screening of a candidate substance that interact with a GENSET polypeptide, fragments or variants thereof, said methods comprising the following steps:

- a) providing a polypeptide comprising, consisting essentially of, or consisting of a GENSET protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide selected from the group consisting of sequences of SEQ ID NOs: 170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool;
- b) obtaining a candidate substance;
- c) bringing into contact said polypeptide with said candidate substance;
- d) detecting the complexes formed between said polypeptide and said candidate substance.

The invention further relates to a method for the production of a pharmaceutical composition comprising a method for the screening of a candidate substance that interact with a GENSET polypeptide, fragments or variants thereof and furthermore mixing the identified substance with a pharmaceutically acceptable carrier.

The invention further concerns a kit for the screening of a candidate substance interacting with the GENSET polypeptide, wherein said kit comprises:

- a) a polypeptide comprising, consisting essentially of, or consisting of a GENSET protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a polypeptide

selected from the group consisting of sequences of SEQ ID NOs:170-338, 456-560, 785-918 and polypeptides encoded by the clone inserts of the deposited clone pool; and

b) optionally means useful to detect the complex formed between said polypeptide or a variant thereof and the candidate substance.

- 5 In a preferred embodiment of the kit described above, the detection means comprises a monoclonal or polyclonal antibody binding to said GENSET protein or fragment or variant thereof.

In vivo methods

- Compounds that suppress or enhance GENSET expression can also be identified using *in vivo* screens. In these assays, the test compound is administered (e.g. IV, IP, IM, orally, or
10 otherwise), to the animal, for example, at a variety of dose levels. The effect of the compound on GENSET expression is determined by comparing GENSET levels, for example in tissues known to express the gene of interest using, for example the data obtained in Tables III, IV, or V, and using Northern blots, immunoassays, PCR, etc., as described above. Suitable test animals include, but are not limited to, rodents (e.g., mice and rats), primates, and rabbits. Humanized mice can also be
15 used as test animals, that is mice in which the endogenous mouse protein is ablated (knocked out) and the homologous human protein added back by standard transgenic approaches. Such mice express only the human form of a protein. Humanized mice expressing only the human GENSET can be used to study *in vivo* responses to potential agents regulating GENSET protein or mRNA levels. As an example, transgenic mice have been produced carrying the human apoE4 gene. They
20 are then bred with a mouse line that lacks endogenous apoE, to produce an animal model carrying human proteins believed to be instrumental in development of Alzheimer's pathology. Such transgenic animals are useful for dissecting the biochemical and physiological steps of disease, and for development of therapies for disease intervention (Loring, *et al*, 1996) (incorporated herein by reference in its entirety).

25 **Uses for compounds modulating GENSET expression and/or biological activity**

- Using *in vivo* (or *in vitro*) systems, it may be possible to identify compounds that exert a tissue specific effect, for example, that increase GENSET expression or activity only in tissues of interest, such as the adrenal gland, bone marrow, brain, cerebellum, colon, fetal brain, fetal kidney, fetal liver, heart, hypertrophic prostate, kidney, liver, lung, lymph ganglia, lymphocytes, muscle,
30 ovary, pancreas, pituitary gland, placenta, prostate, salivary gland, spinal cord, spleen, stomach, intestine, substantia nigra, testis, thyroid, umbilical cord, and uterus. Screening procedures such as those described above are also useful for identifying agents for their potential use in pharmacological intervention strategies. Agents that enhance GENSET gene expression or stimulate its activity may thus be used to induce any phenotype associated with a GENSET gene, or
35 to treat disorders resulting from a deficiency of a GENSET polypeptide activity or expression. Compounds that suppress GENSET polypeptide expression or inhibit its activity can be used to

treat any disease or condition associated with increased or deleterious GENSET polypeptide activity or expression.

Also encompassed by the present invention is an agent which interacts with a GENSET gene or polypeptide directly or indirectly, and inhibits or enhances GENSET polypeptide expression and/or function. In one embodiment, the agent is an inhibitor which interferes with a GENSET polypeptide directly (e.g., by binding the GENSET polypeptide) or indirectly (e.g., by blocking the ability of the GENSET polypeptide to have a GENSET biological activity). In a particular embodiment, an inhibitor of a GENSET protein is an antibody specific for the GENSET protein or a functional portion of the GENSET protein; that is, the antibody binds a GENSET polypeptide. For example, the antibody can be specific for a polypeptide encoded by one of the nucleic acid sequences of human GENSETs (SEQ ID NOs:1-169, 339-455, 561-784), a mammalian GENSET nucleic acid, or portions thereof. Alternatively, the inhibitor can be an agent other than an antibody (e.g., small organic molecule, protein or peptide) which binds the GENSET polypeptide and blocks its activity. For example, the inhibitor can be an agent which mimics the GENSET polypeptide structurally, but lacks its function. Alternatively, it can be an agent which binds to or interacts with a molecule which the GENSET polypeptide normally binds to or interacts with, thus blocking the GENSET polypeptide from doing so and preventing it from exerting the effects it would normally exert.

In another embodiment, the agent is an enhancer (activator) of a GENSET polypeptide which increases the activity of the GENSET polypeptide (increases the effect of a given amount or level of GENSET), increases the length of time it is effective (by preventing its degradation or otherwise prolonging the time during which it is active) or both either directly or indirectly. For example, GENSET polynucleotides and polypeptides can be used to identify drugs which increase or decrease the ability of GENSET polypeptides to induce GENSET biological activity, which drugs are useful for the treatment or prevention of any disease or condition associated with a GENSET biological activity.

The GENSET sequences of the present invention can also be used to generate nonhuman gene knockout animals, such as mice, which lack a GENSET gene or transgenically overexpress a GENSET gene. For example, such GENSET gene knockout mice can be generated and used to obtain further insight into the function of the GENSET gene as well as assess the specificity of GENSET activators and inhibitors. Also, over expression of the GENSET gene (e.g., a human GENSET gene) in transgenic mice can be used as a means of creating a test system for GENSET activators and inhibitors (e.g., against a human GENSET polypeptide). In addition, the GENSET gene can be used to clone the GENSET promoter/enhancer in order to identify regulators of GENSET gene transcription. GENSET gene knockout animals include animals which completely or partially lack the GENSET gene and/or GENSET activity or function. Thus the present invention relates to a method of inhibiting (partially or completely) a GENSET biological activity in a mammal (e.g., a human), the method comprising administering to the mammal an effective amount of an inhibitor of a GENSET polypeptide or polynucleotide. The invention also relates to a

method of enhancing a GENSET biological activity in a mammal, the method comprising administering to the mammal an effective amount of an enhancer of a GENSET polypeptide or polynucleotide.

Inhibiting GENSET gene expression

- 5 Therapeutic compositions according to the present invention may comprise advantageously one or several GENSET oligonucleotide fragments as an antisense tool or a triple helix tool that inhibits the expression of the corresponding GENSET gene.

Antisense Approach

- 10 In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel *et al.*(1995), which disclosure is hereby incorporated by reference in its entirety.

- 15 Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to GENSET mRNA, more preferably to the 5' end of the GENSET mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

- 20 Other preferred antisense polynucleotides according to the present invention are sequences complementary to either a sequence of GENSET mRNAs comprising the translation initiation codon ATG or a sequence of GENSET genomic DNA containing a splicing donor or acceptor site.

- 25 Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu *et al.*(1994), which disclosure is hereby incorporated by reference in its entirety. In a preferred embodiment, these GENSET antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner *et al.*(1991), which disclosure is hereby incorporated by reference in its
30 entirety.

- 35 The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the GENSET mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et al.*, (1986) and Izant and Weintraub, (1984), the disclosures of which are incorporated herein by reference.

 In some strategies, antisense molecules are obtained by reversing the orientation of the GENSET coding region with respect to a promoter so as to transcribe the opposite strand from that

which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of GENSET antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

5 Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use
10 in antisense strategies include 2' O-methyl RNA oligonucleotides and Protein-nucleic acid (PNA) oligonucleotides. Further examples are described by Rossi *et al.*, (1991), which disclosure is hereby incorporated by reference in its entirety.

Various types of antisense oligonucleotides complementary to the sequence of the GENSET cDNA or genomic DNA may be used. In one preferred embodiment, stable and semi-stable
15 antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

20 In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby incorporated by reference, are
25 used. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages, wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic
30 spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, incorporated by reference, may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds
35 to control proteins and are effective as decoys therefor. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2, hereby incorporated by reference are used. These

ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, hereby incorporated by reference, is also contemplated. Because these molecules have
5 no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the
10 cells by diffusion, injection, infection or transfection using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art,
15 including retroviral or viral vectors, vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between $1 \times 10^{-10} \text{M}$ to $1 \times 10^{-4} \text{M}$. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage
20 suitable for use *in vivo*. For example, an inhibiting concentration in culture of 1×10^{-7} translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

25 In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

An alternative to the antisense technology that is used according to the present invention
30 comprises using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme comprises (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage
35 products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as described by Rossi *et al*, (1991) and

Sczakiel *et al.* (1995), the specific preparation procedures being referred to in said articles being herein incorporated by reference.

Triple Helix Approach

The GENSET genomic DNA may also be used to inhibit the expression of the GENSET
5 gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene. The GENSET cDNAs or genomic DNAs of the present invention or, more preferably, a fragment of those sequences, can be used to inhibit gene expression in individuals having diseases
10 associated with expression of a particular gene. Similarly, a portion of the GENSET genomic DNA can be used to study the effect of inhibiting GENSET gene transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both
15 types of sequences from the GENSET genomic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the GENSET genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting GENSET
20 expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting GENSET expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the GENSET gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to
25 those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced GENSET expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the GENSET gene in cells which have been treated with the
30 oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiology within cells derived from individuals with a particular inherited disease, particularly when the cDNA is associated with the disease using
35 techniques described in the section entitled "Identification of genes associated with hereditary diseases or drug response".

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques and at a dosage calculated based on the *in vitro* results, as described in the section entitled "Antisense Approach".

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.* (1989), which is hereby incorporated by this reference.

10 Treating GENSET gene-related disorders

The present invention further relates to methods, uses of GENSET polypeptides and polynucleotides, and uses of modulators of GENSET polypeptides and polynucleotides, for treating diseases/disorders associated with GENSET genes by increasing or decreasing GENSET gene activity and/or expression. These methodologies can be effected using compounds selected using screening protocols such as those described herein and/or by using the gene therapy and antisense approaches described in the art and herein. Gene therapy can be used to effect targeted expression of GENSET genes in any tissue, e.g. a tissue associated with the disease or condition to be treated. The GENSET coding sequence can be cloned into an appropriate expression vector and targeted to a particular cell type(s) to achieve efficient, high level expression. Introduction of the GENSET coding sequence into target cells can be achieved, for example, using particle mediated DNA delivery, (Haynes, 1996 and Maurer, 1999), direct injection of naked DNA, (Levy *et al.*, 1996; and Felgner, 1996), or viral vector mediated transport (Smith *et al.*, 1996, Stone et al, 2000; Wu and Atai, 2000), each of which disclosures are hereby incorporated by reference in their entireties. Tissue specific effects can be achieved, for example, in the case of virus mediated transport by using viral vectors that are tissue specific, or by the use of promoters that are tissue specific. For instance, any tissue-specific promoter may be used to achieve specific expression, for example albumin promoters (liver specific; Pinkert et al., 1987 Genes Dev. 1:268-277), lymphoid specific promoters (Calame et al., 1988 Adv. Immunol. 43:235-275), promoters of T-cell receptors (Winoto et al., 1989 EMBO J. 8:729-733) and immunoglobulins (Banerji et al., 1983 Cell 33:729-740; Queen and Baltimore 1983 Cell 33:741-748), neuron-specific promoters (e.g. the neurofilament promoter; Byrne et al., 1989 Proc. Natl. Acad. Sci. USA 86:5473-5477), pancreas-specific promoters (Edlunch et al., 1985 Science 230:912-916) or mammary gland-specific promoters (milk whey promoter, U.S. Pat. No. 4,873,316 and European Application Publication No. 264, 166). Developmentally-regulated promoters can also be used, such as the murine homeobox promoters (Kessel et al., 1990 Science 249:374-379) or the alpha-fetoprotein promoter (Campes et al., 1989 Genes Dev. 3:537-546).

Combinatorial approaches can also be used to ensure that the GENSET coding sequence is activated in the target tissue (Butt and Karathanasis, 1995; Miller and Whelan, 1997), which

disclosures are hereby incorporated by reference in their entireties. Antisense oligonucleotides complementary to GENSET mRNA can be used to selectively diminish or ablate the expression of the protein, for example, at sites of inflammation. More specifically, antisense constructs or antisense oligonucleotides can be used to inhibit the production of GENSET in high expressing
5 cells such as those cited in the third column of Table V. Antisense mRNA can be produced by transfecting into target cells an expression vector with the GENSET gene sequence, or a portion thereof, oriented in an antisense direction relative to the direction of transcription. Appropriate vectors include viral vectors, including retroviral, adenoviral, and adeno-associated viral vectors, as well as nonviral vectors. Tissue specific promoters can be used, as described supra. Alternatively,
10 antisense oligonucleotides can be introduced directly into target cells to achieve the same goal. (See also other delivery methodologies described herein in connection with gene therapy.). Oligonucleotides can be selected/designed to achieve a high level of specificity (Wagner *et al.*, 1996), which disclosure is hereby incorporated by reference in its entirety. The therapeutic methodologies described herein are applicable to both human and non-human mammals (including
15 cats and dogs).

PHARMACEUTICAL AND PHYSIOLOGICALLY ACCEPTABLE COMPOSITIONS

The present invention also relates to pharmaceutical or physiologically acceptable compositions comprising, as active agent, the polypeptides, nucleic acids or antibodies of the invention. The invention also relates to compositions comprising, as active agent, compounds
20 selected using the above-described screening protocols. Such compositions include the active agent in combination with a pharmaceutical or physiologically acceptable carriers. In the case of naked DNA, the "carrier" may be gold particles. The amount of active agent in the composition can vary with the agent, the patient and the effect sought. Likewise, the dosing regimen can vary depending on the composition and the disease/disorder to be treated.

25 Therefore, the invention related to methods for the production of pharmaceutical composition comprising a method for selecting an active agent, compound, substance or molecule using any of the screening method described herein and furthermore mixing the identified active agent, compound, substance or molecule with a pharmaceutically acceptable carrier.

The pharmaceutical compositions utilized in this invention may be administered by any
30 number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations
35 which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co. Easton, Pa).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through a combination of active compounds with solid excipient, sulting mixture is optionally grinding, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titaniumdioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquidpolyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethylcellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may
5 contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of a GENSET polypeptide, such labeling would include amount, frequency, and method of
10 administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell
15 culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example a
20 GENSET polypeptide or fragments thereof, antibodies specific to GENSET polypeptides, agonists, antagonists or inhibitors of GENSET polypeptides, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between
25 therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within
30 this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account
35 include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions maybe administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

USES OF GENSET SEQUENCES: COMPUTER-RELATED EMBODIMENTS

As used herein the term “cDNA codes of SEQ ID NOs:1-169, 339-455, 561-784” encompasses the nucleotide sequences of SEQ ID NOs:1-169, 339-455, 561-784 and of clones inserts of the deposited clone pool, fragments thereof, nucleotide sequences homologous thereto, and sequences complementary to all of the preceding sequences. The fragments include fragments of SEQ ID NOs:1-169, 339-455, 561-784 comprising at least 8, 10, 12, 15, 18, 20, 25, 28, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, 500, 1000 or 2000 consecutive nucleotides of SEQ ID NOs:1-169, 339-455, 561-784. Preferably the fragments include polynucleotides described herein as encoding polypeptides having a biological activity. Homologous sequences and fragments of SEQ ID NOs:1-169, 339-455, 561-784 refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% identity to these sequences. Identity may be determined using any of the computer programs and parameters described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also include RNA sequences in which uridines replace the thymines in the cDNA codes of SEQ ID NOs:1-169, 339-455, 561-784. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. It will be appreciated that the cDNA codes of SEQ ID NOs:1-169, 339-455, 561-784 can be represented in the traditional single character format (see, e.g. the inside back cover of Stryer, 1995) or in any other format which records the identity of the nucleotides in a sequence.

As used herein the term “polypeptide codes of SEQ ID NOs:170-338, 456-560, 785-918” encompasses the polypeptide sequences of SEQ ID NOs:170-338, 456-560, 785-918 which are encoded by the cDNAs of SEQ ID NOs:1-169, 339-455, 561-784, the polypeptide sequences encoded by the clone inserts of the deposited clone pool, polypeptide sequences homologous thereto, or fragments of any of the preceding sequences. Homologous polypeptide sequences refer to a polypeptide sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% identity to one of the polypeptide sequences of SEQ ID NOs:170-338, 456-560, 785-918. Identity may be determined using any of the computer programs and parameters described herein, including FASTA with the default parameters or with any modified parameters. The homologous sequences may be obtained using any of the procedures described herein or may result from the correction of a sequencing error as described above. The polypeptide fragments comprise at least 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 60, 75, 100, 150 or 200 consecutive amino acids of the polypeptides of

SEQ ID NOs: 170-338, 456-560, 785-918. Preferably, the fragments include polypeptides described herein as having a biological activity, or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of polypeptides described herein as having a biological activity. It will be appreciated that the polypeptide codes of the SEQ ID NOs: 170-338, 456-560, 5 785-918 can be represented in the traditional single character format or three letter format (see, the inside back cover of Stryer, 1995) or in any other format which relates the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium 10 which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the polypeptide codes of the invention. Another aspect of the present 15 invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, 20 electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems 25 which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 1. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the computer system 100 30 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines.

35 Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system 100 further includes one
5 or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc.
10 containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other
15 computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

20 In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of the invention or the polypeptide codes of the invention stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or
25 polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to
30 identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Figure 2 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the
35 homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for
5 analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for
10 comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a
15 determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state
20 notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state
25 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision
30 state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a polypeptide code of the invention, a data storage device having retrievably stored thereon reference
35 nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify motifs implicated in biological function and structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural

motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or polypeptide codes of the invention.

5 Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer
10 program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of the invention through the use of the computer program and
15 determining homology between the nucleic acid codes and reference nucleotide sequences.

Figure 3 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then
20 moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

25 A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250
30 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there are no more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that
35 were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide

sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single base substitution, insertion, or deletion.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of the invention and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above and the method illustrated in Figure 3. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of the invention.

Figure 4 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name

“TAATAA Box” and the feature attribute would be “TAATAA”. An example of such a database is produced by the University of Wisconsin Genetics Computer Group (www.gcg.com).

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether more features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of the invention. Such programs may use any methods known to those skilled in the art including methods based on homology-modeling, fold recognition and *ab initio* methods as described in Sternberg *et al.*, 1999, which disclosure is hereby incorporated by reference in its entirety. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg *et al.*, U.S. Patent No. 5,436,850 issued July 25, 1995, which disclosure is hereby incorporated by reference in its entirety). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of the invention. (See e.g., Srinivasan, *et al.*, U.S. Patent No. 5,557,535 issued September 17, 1996, which disclosure is hereby incorporated by reference in its entirety). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini *et al.*, (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is

performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

5 According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained
10 from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi *et al.*, (1997)).

15 The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of the invention.

 Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of the invention or the polypeptide codes of the invention comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or
20 polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies linear or structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the
25 invention or the polypeptide codes of the invention through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

 The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or
30 WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the invention. The following list is intended not to limit the invention but to provide guidance to
35 programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul *et al.*, 1990),

FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag *et al.*, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwent's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

CONCLUSION

As discussed above, the GENSET polynucleotides and polypeptides of the present invention or fragments thereof can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; as a reagent (including a labeled reagent) in assays designed to quantitatively determine levels of GENSET expression in biological samples; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris *et al.*, (1993) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning; A Laboratory Manual", 2d ed., Cole Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology; Guide to Molecular Cloning Techniques", Academic Press, Berger and Kimmel eds., 1987, which disclosures are hereby incorporated by reference in their entireties.

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims.

EXAMPLES

Preparation of Antibody Compositions to the GENSET protein

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the GENSET protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an

Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the GENSET protein or a portion thereof can be
5 prepared from murine hybridomas according to the classical method of Kohler and Milstein, (1975) or derivative methods thereof. Also see Harlow and Lane. (1988).

Briefly, a mouse is repetitively inoculated with a few micrograms of the GENSET protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol
10 with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall,
15 (1980), which disclosure is hereby incorporated by reference in its entirety, and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, *et al.* (1986) Section 21-2.

20 B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the GENSET protein or a portion thereof can be prepared by immunizing suitable non-human animal with the GENSET protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected,
25 usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for GENSET concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such
30 agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

35 Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng

level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (1971), which disclosure is hereby incorporated by reference in its entirety.

5 Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony *et al.*, (1973), which disclosure is hereby incorporated by reference in its entirety. Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 uM). Affinity of the
10 antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (1980), which disclosure is hereby incorporated by reference in its entirety.

 Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to
15 identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

REFERENCES

- Abbondanzo *et al.*, (1993), Meth. Enzymol., Academic Press, New York, pp 803-823
- 20 Altschul *et al.*, (1990), J. Mol. Biol. 215(3):403-410
- Altschul *et al.*, (1993), Nature Genetics 3:266-272
- Altschul *et al.*, (1997), Nuc. Acids Res. 25:3389-3402
- Ames *et al.*, (1995), J. Immunol. Meth. 184:177-186.
- Anton and Graham, (1995), J. Virol., 69: 4600-4606
- 25 Araki *et al.*, (1995) Proc. Natl. Acad. Sci. U S A. 92(1):160-4.
- Ashkenazi *et al.*, (1991), Proc. Natl. Acad. Sci. U S A 88:10535-10539.
- Aszódi *et al.*, (1997) Proteins: Structure, Function, and Genetics, Supplement 1:38-42
- Attwood *et al.*, (1996) Nucleic Acids Res. 24(1):182-8.
- Attwood *et al.*, (2000) Nucleic Acids Res. 28(1):225-7
- 30 Bartunek *et al.*, (1996), Cytokine. 8(1):14-20.
- Bateman *et al.*, (2000) Nucleic Acids Res. 28(1):263-6
- Baubonis (1993) *Nucleic Acids Res.* 21(9):2025-9.
- Beaucage *et al.*, (1981) *Tetrahedron Lett*, 22: 1859-1862
- Benham *et al.* (1989) *Genomics* 4:509-517,
- 35 Better *et al.*, (1988), *Science*. 240:1041-1043.
- Bittle *et al.*, (1985), *Virol.* 66:2347-2354.
- Bowie *et al.*, (1994), *Science*. 247:1306-1310.

- Bradley (1987), Production and analysis of chimaeric mice. *In*: E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, pp.113.
- Bram *et al.*, (1993), Mol. Cell Biol., 13 : 4760-4769
- Brinkman *et al.*, (1995) J. Immunol Methods. 182:41-50.
- 5 Brown *et al.*, (1979) Meth. Enzymol. 68:109-151
- Brutlag *et al.* (1990) Comp. App. Biosci. 6:237-245
- Bucher and Bairoch (1994) Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman et al, Eds., pp53-61, AAAIPress, Menlo Park.
- Burton *et al.* (1994), Adv. Immunol. 57:191-280
- 10 Bush *et al.*, (1997), J. Chromatogr., 777 : 311-328.
- Butt and Karathanasis (1995) Gene Expr. 4(6):319-36.
- Carlson *et al.*, (1997), J. Biol. Chem. 272(17):11295-11301.
- Chai et al., (1993) Biotechnol. Appl. Biochem. 18:259-273.
- Chang *et al.*, (1993) Gene 127:95-8
- 15 Chee *et al.*, (1996) Science. 274:610-614.
- Chen *et al.* (1987) Mol. Cell. Biol. 7:2745-2752.
- Chen *et al.*, (1998), Cancer Res. 58(16):3668-3678.
- Cherif et al., (1990) Proc. Natl. Acad. Sci. U.S.A., 87:6639-6643
- Cho *et al.*, (1998), Proc. Natl. Acad. Sci. USA, 95(7) : 3752-3757.
- 20 Chou, (1989), Mol. Endocrinol. 3: 1511-1514.
- Chow et al., (1985), Proc. Natl. Acad. Sci. USA. 82:910-914.
- Cleland et al., (1993), Crit. Rev. Therapeutic Drug Carrier Systems. 10:307-377.
- Coles *et al.*, (1998) Hum Mol Genet 7:791-800
- Compton (1991) Nature 350(6313):91-92.
- 25 Corpet *et al.* (2000) Nucleic Acids Res. 28(1):267-9
- Cox *et al.*, (1990) Science 250:245-250
- Creighton (1983), Proteins: Structures and Molecular Principles, W.H. Freeman & Co. 2nd Ed., T. E., New York
- Creighton, (1993) , Posttranslational Covalent Modification of Proteins, W.H. Freeman and
- 30 Company, New York B.C. Johnson, Ed., Academic Press, New York 1-12
- Cunningham *et al.* (1989), Science 244:1081-1085.
- Davis *et al.*, (1986) Basic Methods in Molecular Biology, ed., Elsevier Press, NY,
- Decker and Parker, (1995) Curr. Opin. Cell. Biol. 7(3) :368-92
- Dempster *et al.*, (1977) Stat. Soc., 39B:1-38.
- 35 Deng *et al.*, (1998) Blood. 92(6):1981-1988.
- Dent and Latchman (1993) The DNA mobility shift assay. In: Transcription Factors: A Practical Approach (Latchman DS, ed.) pp1-26. Oxford: IRL Press

- Derrigo *et al.*, (2000) Int. J. Mol. Med. 5(2) :111-23
- Eckner *et al.*, (1991) EMBO J. 10:3513-3522.
- Edwards and Leatherbarrow, (1997) Analytical Biochemistry, 246, 1-6
- Engvall, (1980) Meth. Enzymol. 70:419
- 5 Erlich, (1992) PCR Technology; Principles and Applications for DNA Amplification. W.H. Freeman and Co., New York
- Feldman and Steg, (1996), Medecine/Sciences, 12:47-55
- Felgner (1996) Hum Gene Ther. 7(15):1791-3.
- Felici, (1991), J. Mol. Biol., 222:301-310
- 10 Fell *et al.*, (1991), J. Immunol. 146:2446-2452.
- Fields and Song, (1989), Nature, 340: 245-246
- Fisher, (1980) Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C.
- Flotte *et al.*, (1992) Am. J. Respir. Cell Mol. Biol. 7:349-356.
- 15 Fodor *et al.*, (1991) Science 251:767-777.
- Foster *et al.*, (1996) Genomics 33:185-192
- Fountoulakis *et al.*, (1995) Biochem. 270:3958-3964.
- Fraley *et al.*, (1979) Proc. Natl. Acad. Sci. USA. 76:3348-3352.
- Frazer *et al.*, (1992) Genomics 14:574-584
- 20 Fried and Crothers, (1981) Nucleic Acids Res. 9:6505-6525
- Fromont-Racine *et al.*, (1997), Nature Genetics, 16(3) : 277-282.
- Fry *et al.*, (1992) Biotechniques, 13: 124-131
- Fudenberg, (1980) Chap. 26 in: Basic 503 Clinical Immunology, 3rd Ed. Lange, Los Altos, California
- 25 Fuller S. A. *et al.* (1996) Immunology in Current Protocols in Molecular Biology, Furth P.A. *et al.* (1994) Proc. Natl. Acad. Sci USA. 91:9302-9306.
- Garner and Revzin, (1981) Nucleic Acids Res 9:3047-3060
- Gentz *et al.*, (1989) Proc Natl Acad Sci U S A. 86(3):821-4.
- Geysen *et al.*, (1984), Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002.
- 30 Ghosh and Bacchawat, (1991), Targeting of liposomes to hepatocytes, IN: Liver Diseases, Targeted diagnosis and therapy using specific receptors and ligands. Eds., Marcel Dekker, New York, pp. 87-104.
- Gillies *et al.*, (1989), J. Immunol Methods. 125:191-202.
- Gillies *et al.*, (1992), Proc Natl Acad Sci U S A 89:1428-1432.
- 35 Gonnet *et al.*, (1992), Science 256:1443-1445
- Gopal (1985) Mol. Cell. Biol., 5:1188-1190.
- Gossen *et al.*, (1992) Proc. Natl. Acad. Sci. USA. 89:5547-5551.

- Gossen *et al.*, (1995) *Science*. 268:1766-1769.
- Graham *et al.*, (1973) *Virol.* 52:456-457.
- Green *et al.*, (1986) *Ann. Rev. Biochem.* 55:569-597
- Greenspan and Bona (1989), *FASEB J.* 7(5):437-444.
- 5 Griffais *et al.*, (1991) *Nucleic Acids Res.* 19: 3887-3891
- Griffin *et al.*, (1989) *Science* 245:967-971
- Gu H. *et al.*, (1993) *Cell* 73:1155-1164.
- Gu H. *et al.*, (1994) *Science* 265:103-106.
- Guatelli *et al.*, (1990) *Proc. Natl. Acad. Sci. USA.* 35:273-286.
- 10 Gyuris *et al.*, (1993) *Cell* 75:791-803
- Hames and Higgins (1985) *Nucleic Acid Hybridization: A Practical Approach*. Hames and Higgins Ed., IRL Press, Oxford.
- Hammerling (1981), *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y. 563-681.
- 15 Hansson *et al.*, (1999), *J. Mol. Biol.* 287:265-276.
- Haravama (1998), *Trends Biotechnol.* 16(2): 76-82.
- Harland *et al.*, (1985) *J. Cell. Biol.* 101:1094-1095.
- Harlow and Lane, (1988) *Antibodies A Laboratory Manual*. Cold Spring Harbor Laboratory. pp. 53-242
- 20 Harper *et al.*, (1993), *Cell*, 75 : 805-816
- Harrop *et al.*, (1998), *J. Immunol.* 161(4):1786-1794.
- Haynes *et al.*, (1996) *J Biotechnol.* 44(1-3):37-42.
- Henikoff and Henikoff, (1993), *Proteins* 17:49-61
- Henikoff *et al.*, (2000) *Electrophoresis* 21(9):1700-6
- 25 Henikoff *et al.*, (2000) *Nucleic Acids Res.* 28(1):228-30
- Higgins *et al.*, (1996), *Meth. Enzymol.* 266:383-402
- Hillier and Green (1991) *PCR Methods Appl.*, 1: 124-8.
- Hoess *et al.*, (1986) *Nucleic Acids Res.* 14:2287-2300.
- Hofmann *et al.*, (1999) *Nucl. Acids Res.* 27:215-219.;
- 30 Holm and Sander (1996) *Nucleic Acids Res.* 24(1):206-9
- Holm and Sander (1997) *Nucleic Acids Res.* 25(1):231-4
- Holm and Sander (1999) *Nucleic Acids Res.* 27(1):244-7
- Hoppe *et al.*, (1994), *FEBS Letters.* 344:191.
- Houghten (1985), *Proc. Natl. Acad. Sci. USA* 82:5131-5135.
- 35 Huang *et al.*, (1996) *Cancer Res* 56(5):1137-1141.
- Hunkapiller *et al.*, (1984) *Nature.* 310(5973):105-11.
- Huston *et al.*, (1991), *Meth. Enymol.* 203:46_88.

- Huygen *et al.*, (1996) *Nature Medicine*. 2(8):893-898.
- Izant and Weintraub, (1984) *Cell* 36(4):1007-15
- Jameson and Wolf, (1988), *Comp. Appl. Biosci.* 4:181-186
- Julan *et al.*, (1992) *J. Gen. Virol.* 73:3251-3255.
- 5 Kanegae *et al.*, (1995) *Nucl. Acids Res.* 23:3816-3821.
- Karlin and Altschul, (1990), *Proc. Natl. Acad. Sci. USA* 87:2267-2268
- Kettleborough *et al.*, (1994), *Eur. J Immunol.* 24:952-958.
- Kim U-J. *et al.*, (1996) *Genomics* 34:213-218.
- Klein *et al.*, (1987) *Nature*. 327:70-73.
- 10 Kohler and Milstein, (1975) *Nature* 256:495
- Koller *et al.*; (1992) *Annu. Rev. Immunol.* 10:705-730.
- Kostelny *et al.*, (1992), *J. Immunol.* 148:1547-1553.
- Landschulz *et al.*, (1988), *Science*. 240:1759.
- Ledbetter *et al.*, (1990) *Genomics* 6:475-481
- 15 Lenhard *et al.*, (1996) *Gene*. 169:187-190.
- Levy *et al.*, (1996) *Gene Ther.* 3(3):201-11.
- Lewin, (1989), *Proc. Natl. Acad. Sci. USA* 86:9832-8935.
- Liautard *et al.*, (1997), *Cytokine*. 9(4):233-241.
- Linton *et al.*, (1993) *J. Clin. Invest.* 92:3029-3037.
- 20 Liu *et al.*, (1994) *Proc. Natl. Acad. Sci. USA*. 91: 4528-4262.
- Lo Conte *et al.*, (2000) *Nucleic Acids Res.* 28(1):257-9.
- Lockhart *et al.*, (1996) *Nature Biotechnology* 14: 1675-1680
- Lorenzo and Blasco (1998) *Biotechniques*. 24(2):308-313.
- Lucas (1994), In : *Development and Clinical Uses of Haemophilus b Conjugate*;
- 25 Makrides, (1999) *Protein Expr. Purif.* 17(2) :183-202
- Malik *et al.*, (1992), *Exp. Hematol.* 20:1028-1035.
- Mansour *et al.*, (1988) *Nature*. 336:348-352.
- Marshall *et al.*, (1994) *PCR Methods and Applications*. 4:80-84.
- Maurer *et al.*, (1999) *Mol Membr Biol.* 16(1):129-40.
- 30 McCormick *et al.*, (1994) *Genet. Anal. Tech. Appl.* 11:158-164.
- McLaughlin *et al.*, (1996) *Am. J. Hum. Genet.* 59:561-569.
- Miller and Whelan, (1997) *Hum Gene Ther.* 8(7):803-15.
- Muller *et al.*, (1998), *Structure*. 6(9):1153-1167.
- Mullinax *et al.*, (1992), *BioTechniques*. 12(6):864-869.
- 35 Murvai *et al.*, (2000) *Nucleic Acids Res.* 28(1):260-2
- Murzin *et al.*, (1995) *J Mol Biol.* 247(4):536-40
- Muzyczka *et al.*, (1992) *Curr. Topics in Micro. and Immunol.* 158:97-129.

- Nada *et al.*, (1993) Cell 73:1125-1135.
- Nagaraja *et al.*, (1997) Genome Research 7:210-222
- Nagy *et al.*, (1993), Proc. Natl. Acad. Sci. USA 90: 8424-8428.
- Nakai and Horton, (1999) Trends Biochem. Sci., 24:34-36
- 5 Nakai and Kanehisa (1992) Genomics 14, 897-911
- Naramura *et al.*, (1994), Immunol. Lett. 39:91-99.
- Narang *et al.*, (1979), Methods Enzymol 68:90-98
- Neda *et al.*, (1991) J. Biol. Chem. 266:14143-14146.
- Nevill-Manning *et al.*, (1998) Proc. Natl. Acad. Sci. U S A. 95, 5865-5871
- 10 Nicolau *et al.*, (1982) Biochim. Biophys. Acta. 721:185-190.
- Nicolau *et al.*, (1987), Meth. Enzymol., 149:157-76.
- Nissinoff, (1991), J. Immunol. 147(8): 2429-2438.
- O'Reilly *et al.*, (1992) Baculovirus Expression Vectors: A Laboratory Manual. W. H. Freeman and Co., New York.
- 15 Obermayr *et al.*, (1996) Eur. J. Hum. Genet. 4:242-245
- Ohno *et al.*, (1994) Science. 265:781-784.
- Oi *et al.*, (1986), BioTechniques 4:214.
- Oldenburg *et al.*, (1992), Proc. Natl. Acad. Sci. USA 89:5393-5397.
- Orengo *et al.*, (1997) Structure. 5(8):1093-108
- 20 Ouchterlony *et al.*, (1973) Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell
- Padlan, (1991), Molec. Immunol. 28(4/5):489-498.
- Parmley and Smith, (1988) Gene 73:305-318
- Patten, *et al.* (1997), Curr Opinion Biotechnol. 8:724-733.
- 25 Pearl *et al.*, (2000) Biochem Soc Trans. 28(2):269-75
- Pearson and Lipman, (1988), Proc. Natl. Acad. Sci. USA 85(8):2444-2448
- Pease and William, (1990), Exp. Cell. Res. 190: 209-211.
- Persic *et al.*, (1997), Gene. 1879-81
- Pesole *et al.*, (2000) Nucleic Acids Res, 28(1):193-196
- 30 Peterson *et al.*, (1993), Proc. Natl. Acad. Sci. USA, 90 : 7593-7597.
- Pietu *et al.*, (1996) Genome Research 6:492-503
- Pinckard *et al.*, (1967), Clin. Exp. Immunol 2:331-340.
- Pitard *et al.*, (1997), J. Immunol. Methods. 205(2):177-190.
- Pongor *et al.* (1993) Protein Eng. 6(4):391-5
- 35 Potter *et al.*, (1984) Proc. Natl. Acad. Sci. U.S.A. 81(22):7161-7165.
- Prat *et al.*, (1998), J. Cell. Sci. 111(Pt2):237-247.
- Raeymaekers *et al.*, (1995) Genomics 29:170-178

- Ramunsen *et al.*, (1997), Electrophoresis, 18 : 588-598.
- Rattan *et al.*, (1992) Ann NY Acad Sci 663:48-62
- Reid *et al.*, (1990) Proc. Natl. Acad. Sci. U.S.A. 87:4299-4303.
- Robbins *et al.*, (1987), Diabetes. 36:838-845.
- 5 Robertson, (1987), Embryo-derived stem cell lines. In: E.J. Robertson Ed. Teratocarcinomas and embrionic stem cells: a practical approach. IRL Press, Oxford, pp. 71.
- Roguska *et al.*, (1994), Proc. Natl. Acad. Sci. U.S.A. 91:969-973.
- Ron *et al.*, (1993), Biol Chem., 268 2984-2988.
- Rose *et al.*, (1980) Chap. 12 in: Methods in Immunodiagnosis, 2d Ed. John Wiley 503 Sons, 10 New York
- Rossi *et al.*, (1991) Pharmacol. Ther. 50:245-254,
- Roth *et al.*, (1996) Nature Medicine. 2(9):985-991.
- Roux *et al.*, (1989) Proc. Natl. Acad. Sci. U.S.A. 86:9079-9083.
- Sambrook *et al.*, (1989) Molecular Cloning: A Laboratory Manual. 2ed. Cold Spring 15 Harbor Laboratory, Cold Spring Harbor, New York.
- Samson *et al.*, (1996) Nature, 382(6593):722-725.
- Samulski *et al.*, (1989) J. Virol. 63:3822-3828.
- Sanchez-Pescador (1988) J. Clin. Microbiol. 26(10):1934-1938.
- Sander and Schneider (1991) Proteins. 9(1):56-68.)
- 20 Sauer *et al.*, (1988) Proc. Natl. Acad. Sci. U.S.A. 85:5166-5170.
- Sawai *et al.*, (1995), AJRI 34:26-34.
- Schedl *et al.*, (1993b), Nucleic Acids Res., 21: 4783-4787.
- Schedl *et al.*, 1(993a), Nature, 362: 258-261.
- Schena *et al.* (1995) Science 270:467-470
- 25 Schena *et al.*, (1996), Proc Natl Acad Sci U S A, 93(20):10614-10619.
- Schuler *et al.*, (1996) Science 274:540-546
- Schultz *et al.*, (1998) Proc Natl Acad Sci U S A 95, 5857-5864
- Schwartz and Dayhoff, (1978), eds., Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation
- 30 Sczakiel *et al.*, (1995) Trends Microbiol. 3(6):213-217.
- Seifter *et al.*, (1990) Meth Enzymol 182:626-646
- Shay *et al.*, (1991), Biochem. Biophys. Acta, 1072: 1-7.
- Shizuya *et al.*, (1992) Proc. Natl. Acad. Sci. U.S.A. 89:8794-8797.
- Shu *et al.*, (1993), Proc. Natl. Acad. Sci. U.S.A. 90:7995-7999.
- 35 Skerra *et al.*, (1988), Science 240:1038-1040.
- Smith and Johnson (1988) Gene. 67(1):31-40.
- Smith *et al.*, (1983) Mol. Cell. Biol. 3:2156-2165.

- Smith *et al.*, (1996) Antiviral Res. 32(2):99-115.
- Sonnhammer and Kahn D (1994) Protein Sci. 3(3):482-92
- Sonnhammer *et al.*, (1997) Proteins. 28(3):405-20
- Sosnowski, *et al.*, (1997) Proc Natl Acad Sci U S A 94:1119-1123
- 5 Sowdhamini *et al.*, (1997) Protein Engineering 10:207, 215
- Sternberg (1994) Mamm. Genome. 5:397-404.
- Sternberg (1992) Trends Genet. 8:1-16.
- Sternberg *et al.*, (1999) Curr Opin Struct Biol. 9(3):368-73.
- Stone *et al.*, (2000) J Endocrinol. 164(2):103-18.
- 10 Stryer, (1995) Biochemistry, 4th edition
- Studnicka *et al.*, (1994), Protein Engineering. 7(6):805-814.
- Sutcliffe *et al.*, (1983), Science. 219:660-666.
- Szabo *et al.*, (1995) Curr Opin Struct Biol 5, 699-705
- Tascon *et al.*, (1996) Nature Medicine. 2(8):888-892.
- 15 Taryman *et al.*, (1995), Neuron. 14(4):755-762.
- Tatusov *et al.*, (1997) Science, 278, 631 :637
- Tatusov *et al.*, (2000) Nucleic Acids Res. 28(1):33-6.)
- Te Riele *et al.*, (1990) Nature. 348:649-651.
- Thomas *et al.*, (1986) Cell. 44:419-428.
- 20 Thomas *et al.*, (1987) Cell. 51:503-512.
- Thompson *et al.*, (1994), Nucleic Acids Res. 22(2):4673-4680
- Trauneker *et al.*, (1988), Nature. 331:84-86.
- Tur-Kaspa *et al.*, (1986) Mol. Cell. Biol. 6:716-718.
- Tutt *et al.*, (1991), J. Immunol. 147:60-69.
- 25 Urdea (1988) Nucleic Acids Research. 11:4937-4957.
- Urdea *et al.*, (1991) Nucleic Acids Symp. Ser. 24:197-200.
- Vaitukaitis *et al.*, (1971) J. Clin. Endocrinol. Metab. 33:988-991
- Valadon *et al.*, (1996), J. Mol. Biol., 261:11-22.
- Van der Lugt *et al.*, (1991) Gene. 105:263-267.
- 30 Vil *et al.*, (1992) Proc Natl Acad Sci U S 89:11337-11341.
- Vlasak *et al.*, (1983) Eur. J. Biochem. 135:123-126.
- Wabiko *et al.*, (1986) DNA. 5(4):305-314.
- Wagner *et al.*, (1996) Nat Biotechnol. 14(7):840-4.
- Walker *et al.*, (1996) Clin. Chem. 42:9-13.
- 35 Wang *et al.*, (1997), Chromatographia, 44 : 205-208.
- Warrington *et al.*, (1991) Genomics 11:701-708
- Westerink, (1995), Proc. Natl. Acad. Sci USA., 92:4021-4025

- White (1997) B.A. Ed. in Methods in Molecular Biology 67: Humana Press, Totowa
- White *et al.* (1997) Genomics. 12:301-306.
- Wilson *et al.*, (1984) Cell. 37(3):767-78.
- Wong *et al.*, (1980) Gene. 10:87-94.
- 5 Wood *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82(6):1585-1588
- Wood *et al.*, (1993), Proc. Natl. Acad. Sci. USA, 90: 4582-4585.
- Wu and Atai, (2000) Curr Opin Biotechnol. 11(2):205-8.
- Wu and Wu, (1987) J. Biol. Chem. 262:4429-4432.
- Wu and Wu, (1988) Biochemistry. 27:887-892.
- 10 Yagi T. *et al.*, (1990) Proc. Natl. Acad. Sci. U.S.A. 87:9918-9922.
- Yona *et al.*, (1999) Proteins. 37(3):360-78
- Yoon et al, (1998), J. Immunol. 160(7):3170-3179.
- Zheng, X.X. *et al.* (1995), J. Immunol. 154:5590-5600.
- Zhu *et al.*, (1998), Cancer Res. 58(15):3209-3214.
- 15 Zou *et al.*, (1994) Curr. Biol. 4:1099-1103.

Throughout this application, various publications, patents and published patent applications are cited. The disclosures of these publications, patents and published patent specification referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

TABLE I

SEQ ID NO from priority application (nucl)	SEQ ID NO in present application (nucl.)	SEQ ID NO in present application (prt)	SEQ ID NO in present application (nucl.)	Clone ID
37			NUC561	486589;523280
51			NUC562	500707076;609080;642931
179			NUC563	147941;153834;193100
180			NUC564	147941
183			NUC565	100038;100419;100523;100546
201			NUC566	528046
326			NUC567	484503
362			NUC568	211034;211122
440			NUC569	627628
452			NUC570	500713596;500733538;500741977
483			NUC571	500730326
500			NUC572	482482
505			NUC573	611492
528			NUC574	221311
573			NUC575	144150
574			NUC576	626500
587			NUC577	490129
588			NUC578	129471
593	NUC86	PRT255	NUC406	500762786
599			NUC579	148991;206343;211039
603			NUC580	212418
621			NUC581	395370
628			NUC582	116680
653			NUC583	224425
670			NUC584	225626
678	NUC87	PRT256	NUC407	822794
678	NUC88	PRT257	NUC407	337572
693			NUC585	500760143
703			NUC586	125325;145574;158295;225872;334779
746			NUC587	620699
770			NUC588	500735221
775			NUC589	131662;131668;177901;200591
796			NUC590	500756189
812			NUC591	500694179
940			NUC592	158339;213121;220652;236981;239275;239598;244360;582565
988			NUC593	238123;239495;334569;582920
996			NUC594	334818
1036			NUC595	483794;519407;633595;633902
1064			NUC596	608607
1151			NUC597	500743552;500744660
1190			NUC598	101006;509431
1458			NUC599	313060;313135;313174
1590			NUC600	178255
1853			NUC601	114927
1904			NUC602	107824
2028			NUC603	170306
2173	NUC89	PRT258	NUC408	642374
2368			NUC604	210539;331006

2553			NUC605	106061
2556			NUC606	106061
2658			NUC607	654607
2690			NUC608	172048;172057
2755			NUC609	101090
2800			NUC610	119222;119491;151036
2843			NUC611	619446;619452;633712
2852			NUC612	145573
2932			NUC613	650981
2955			NUC614	130068
3078			NUC615	100966;99482
3280			NUC616	147041
3326			NUC617	132294;132317
3387			NUC618	126303;134231;135124
3439			NUC619	502644
3501			NUC620	625238;632902;635258
3633			NUC621	120631
3678			NUC622	200451
3714	NUC90	PRT259	NUC409	231569
3714	NUC91	PRT260	NUC409	145151
3714			NUC409	153628
3796			NUC623	131690;588080
3801			NUC624	199999
3804			NUC625	483173;510100;650405
3892			NUC626	626911;627852;631702;633566
3985			NUC627	229507;229539;236257
4005			NUC628	237312;335729;490055
4063			NUC629	521127
4088	NUC92	PRT261	NUC410	128061
4088	NUC93	PRT262	NUC410	118027
4088			NUC410	166676
4111	NUC94	PRT263	NUC411	627202
4111			NUC411	538182;620818;625154;628241; 629431;633031;634788
4126			NUC630	153486
4172			NUC631	241664
4261	NUC95	PRT264	NUC412	112311
4340			NUC632	500703884;633346;634598
4436			NUC633	204316
4609	NUC1	PRT170	NUC339	502084
4647	NUC2	PRT171	NUC340	589115
4660			NUC634	643537
4664	NUC3	PRT172	NUC341	1000902917
4678	NUC4	PRT173	NUC342	602517
4678	NUC5	PRT174	NUC342	478210
4678			NUC342	763189
4682	NUC6	PRT175	NUC343	500698315
4687	NUC7	PRT176	NUC344	114180
4687			NUC344	114106;175654;654896
4690			NUC635	114106;211516;654896
4694	NUC8	PRT177	NUC345	338112
4696	NUC9	PRT178	NUC346	338100
4733	NUC10	PRT179	NUC347	784093
4807	NUC11	PRT180	NUC348	1000943975
4809	NUC12	PRT181	NUC349	1000771934
4830	NUC13	PRT182	NUC350	186661
4855	NUC14	PRT183	NUC351	105855

4900	NUC15	PRT184	NUC352	500742698
4908			NUC636	538424
4943			NUC637	144901;170048;206458;240439; 241215
4947	NUC16	PRT185	NUC353	201980
4947	NUC17	PRT186	NUC353	198002
4976			NUC638	248323;248866
5000	NUC18	PRT187	NUC354	500739047
5002	NUC19	PRT188	NUC355	1000904024
5005			NUC639	155986;222313;237229
5011	NUC20	PRT189	NUC356	125817
5040			NUC640	646668
5058			NUC641	224715
5071	NUC21	PRT190	NUC357	147648
5089	NUC22	PRT191	NUC358	1000839315
5117	NUC23	PRT192	NUC359	122473
5141	NUC24	PRT193	NUC360	585770
5141	NUC25	PRT194	NUC360	123996
5162	NUC26	PRT195	NUC361	1000904064
5167	NUC27	PRT196	NUC362	482181
5178	NUC28	PRT197	NUC363	500731597
5192	NUC29	PRT198	NUC364	581232
5214			NUC642	394359
5230	NUC30	PRT199	NUC365	613647
5240	NUC31	PRT200	NUC366	715437
5250	NUC32	PRT201	NUC367	1000878517
5262	NUC33	PRT202	NUC368	544474
5270	NUC34	PRT203	NUC369	143880
5278	NUC35	PRT204	NUC370	1000853793
5358	NUC36	PRT205	NUC371	500732568
5453	NUC37	PRT206	NUC372	427150
5453	NUC38	PRT207	NUC372	593306
5453	NUC39	PRT208	NUC372	593993
5453	NUC40	PRT209	NUC372	590939
5453			NUC372	432874;435627
5494	NUC41	PRT210	NUC373	155600
5494	NUC42	PRT211	NUC373	641537
5499			NUC643	500702809
5533	NUC43	PRT212	NUC374	1000872335
5563	NUC44	PRT213	NUC375	1000852500
5609	NUC45	PRT214	NUC376	500720555
5657	NUC46	PRT215	NUC377	500715373
5691	NUC47	PRT216	NUC378	167435
5748	NUC48	PRT217	NUC379	620429
5748	NUC49	PRT218	NUC379	613335
5806	NUC50	PRT219	NUC380	589848
5806	NUC51	PRT220	NUC380	211883
5806	NUC52	PRT221	NUC380	642603
5806	NUC53	PRT222	NUC380	193316
5816	NUC54	PRT223	NUC381	495917
5824	NUC55	PRT224	NUC382	160935
5861	NUC56	PRT225	NUC383	593736
5885	NUC57	PRT226	NUC384	613887
5913	NUC58	PRT227	NUC385	166601
5947	NUC96	PRT265	NUC413	654627
5966	NUC59	PRT228	NUC386	500762665
5966	NUC60	PRT229	NUC386	500742089

5966	NUC61	PRT230	NUC386	500759088
5970			NUC644	193675;423656
5974	NUC62	PRT231	NUC387	650666
5983			NUC645	626803
5985	NUC63	PRT232	NUC388	594066
6011	NUC97	PRT266	NUC414	1000886279
6080	NUC64	PRT233	NUC389	642569
6081	NUC65	PRT234	NUC390	519656
6108			NUC646	145580
6159	NUC66	PRT235	NUC391	1000903258
6231	NUC67	PRT236	NUC392	715579
6238			NUC647	500694849;500699591;500706028;500710562;500724984;625642;628058;633030
6252	NUC98	PRT267	NUC415	1000855876
6283	NUC68	PRT237	NUC393	820495
6290	NUC69	PRT238	NUC394	500709853
6290	NUC70	PRT239	NUC394	500757399
6290	NUC71	PRT240	NUC394	592868
6322	NUC72	PRT241	NUC395	500739746
6329			NUC648	615173
6334			NUC649	237324
6345	NUC73	PRT242	NUC396	500714172
6345	NUC74	PRT243	NUC396	500716683
6350	NUC75	PRT244	NUC397	1000869553
6358	NUC76	PRT245	NUC398	608537
6384	NUC77	PRT246	NUC399	1000906334
6400			NUC650	149691
6418			NUC651	237026
6431	NUC78	PRT247	NUC400	614334
6453	NUC99	PRT268	NUC416	211056
6636			NUC652	608607
6660			NUC653	129407
6688	NUC100	PRT269	NUC417	646099
6727	NUC79	PRT248	NUC401	199782
6727	NUC80	PRT249	NUC401	821212
6727	NUC81	PRT250	NUC401	202863
6835	NUC101	PRT270	NUC418	158243
6865			NUC654	612052
6892	NUC102	PRT271	NUC419	153261
6892	NUC103	PRT272	NUC419	650872
6892	NUC104	PRT273	NUC419	599054
6892	NUC105	PRT274	NUC419	152042
6892	NUC106	PRT275	NUC419	493328
7000	NUC107	PRT276	NUC420	538694
7041			NUC655	142587; 145561; 146609; 149065; 153394; 153773; 205319; 206906; 215376; 227424; 228016; 240538; 242510; 530873; 588304
7533			NUC656	500758154
7535			NUC657	632835
7577			NUC658	205411
7697	NUC108	PRT277	NUC421	653966
7712	NUC109	PRT278	NUC422	237552
7712			NUC422	202997;206456
8009	NUC110	PRT279	NUC423	645452
8078	NUC111	PRT280	NUC424	335367

8078	NUC112	PRT281	NUC424	334488
8078	NUC113	PRT282	NUC424	329736
8078	NUC114	PRT283	NUC424	244355
8078	NUC115	PRT284	NUC424	150197
8078	NUC116	PRT285	NUC424	244242
8078	NUC117	PRT286	NUC424	223147
8078	NUC118	PRT287	NUC424	221735
8078	NUC119	PRT288	NUC424	215414
8078	NUC120	PRT289	NUC424	149875
8078	NUC121	PRT290	NUC424	167198
8078	NUC122	PRT291	NUC424	193511
8078	NUC123	PRT292	NUC424	226917
8078	NUC124	PRT293	NUC424	225461
8078	NUC125	PRT294	NUC424	193742
8078			NUC424	165071;165245;200864;221825; 243230;581542
8079			NUC659	628867
8097			NUC660	486772;511180
8166	NUC126	PRT295	NUC425	642948
8166	NUC127	PRT296	NUC425	638743
8262			NUC661	151662
8341			NUC662	101420
8534			NUC663	131658;196152;243686
8666	NUC128	PRT297	NUC426	763024
8666	NUC129	PRT298	NUC426	500720430
8671			NUC664	193411
8744			NUC665	162906
8968	NUC82	PRT251	NUC402	771827
8968	NUC130	PRT299	NUC402	500695719
8968			NUC402	620376;635045
8994			NUC666	199362;227277;242546
9297			NUC667	651871
9327			NUC668	247810
9332			NUC669	199155;200810;336623
9406			NUC670	106061
9407			NUC671	106061
9668			NUC672	168218;197771;205623;228775; 238794
9679	NUC131	PRT300	NUC427	206381
9755			NUC673	197091
9868			NUC674	144783;206407;215714;234057; 336758;582582
10044			NUC675	107768;111854;500721812;500 723626;500723636;500724389; 500725580;500729834;5007354 42;500735787;500758255;5007 62395;586703;589397;612312;6 35730;642849;645812;762987;7 67609
10322	NUC132	PRT301	NUC428	200895
10526	NUC133	PRT302	NUC429	1000891255
10584			NUC676	500745219
10650			NUC677	187889;242499
10739	NUC134	PRT303	NUC430	637548
10743	NUC135	PRT304	NUC431	767426
10744	NUC136	PRT305	NUC432	500691428
10761			NUC678	131060
10880	NUC137	PRT306	NUC433	116153

10942	NUC138	PRT307	NUC434	500699885
10942	NUC139	PRT308	NUC434	746303
10942	NUC140	PRT309	NUC434	500705937
10942			NUC434	500705002;500712632;633931; 634489;813634;816859
11019	NUC141	PRT310	NUC435	150568
11278	NUC142	PRT311	NUC436	495638
11342	NUC143	PRT312	NUC437	143196
11562			NUC679	187543
11688			NUC680	165419;165544;166387;181924; 181930;196904;199001;199269; 224447;238731;243770
11735	NUC144	PRT313	NUC438	633418
11735	NUC145	PRT314	NUC438	422878
11735			NUC438	500706283;500711792;5007127 11;500725618;500738973;6513 70
11813			NUC681	633791
12039	NUC146	PRT315	NUC439	546312
12043			NUC682	632330
12048			NUC683	101164
12098			NUC684	135037
12202			NUC685	168232;243338
12220			NUC686	433866
12243			NUC687	659527
12263			NUC688	139596
12276			NUC689	624892;628879;631590;633480
12490			NUC690	500704734;500744586;5007449 72;611533;634337
12604	NUC147	PRT316	NUC440	614106
12604			NUC440	178304
12657			NUC691	238579;248948;397864;521873; 526295;589951
12788	NUC148	PRT317	NUC441	330777
12901	NUC149	PRT318	NUC442	124608
12907	NUC150	PRT319	NUC443	478617
12907	NUC151	PRT320	NUC443	481184
13013			NUC692	583731;650307;650848
13202			NUC693	238886
13229	NUC152	PRT321	NUC444	612301
13256	NUC153	PRT322	NUC445	165123
13256	NUC154	PRT323	NUC445	165643
13256			NUC445	142964;150214;223536;245008
13267	NUC155	PRT324	NUC446	488818
13285			NUC694	193487;238191;248387
26638			NUC695	645819
26710			NUC696	600909;608784;611758;614721; 619435;620041;620372;625815; 625933;625983;626308;627299; 627481;628147;628753;631655; 633039;633371;633760;634553; 642966
26726			NUC697	421115
26786			NUC698	500702480
26982			NUC699	638872
27084	NUC156	PRT325	NUC447	242080
27084			NUC447	128161;186671;210505;211578; 214909;221663;222101;223000; 224361;226849;242326;242424; 243662;244913;247912

27273			NUC700	525674;601556
27301			NUC701	135037
27336			NUC702	500742735
27361			NUC703	205346;530902
27374			NUC704	643006
27627			NUC705	129706;223196
27697			NUC706	500701900
27877			NUC707	99497
28413			NUC708	135042
28517			NUC709	150011;201848
28518			NUC710	500721700;500729093;500730152
29120	NUC157	PRT326	NUC448	488444
29469			NUC711	638852
29472			NUC712	637812
29557			NUC713	188208
29673			NUC714	650606
29814			NUC715	813496
30218			NUC716	241681;242553;589203
30446			NUC717	105288
30477			NUC718	500724995;500758517
30583			NUC719	117932;194613;225013;331614
30719			NUC720	637363
31356			NUC721	106998
31422	NUC158	PRT327	NUC449	500732587
31554			NUC722	222161
31627			NUC723	173050
31726			NUC724	393750
31744			NUC725	176380
31790			NUC726	625728
32102			NUC727	500762549
32473	NUC159	PRT328	NUC450	183902
32475	NUC160	PRT329	NUC451	635993
32962			NUC728	500740719
33130			NUC729	165852;165888
33712	NUC161	PRT330	NUC452	398703
35005			NUC730	124493
35185	NUC83	PRT252	NUC403	589785
35258			NUC731	224898
35326			NUC732	145027
35597			NUC733	237630
35912			NUC734	637431
35984			NUC735	117238
36122			NUC736	226039
37337			NUC737	143508;196052;221995
38112			NUC738	129444
38220			NUC739	608709
38311			NUC740	600921
38631			NUC741	194909
38749			NUC742	163588
38890	NUC162	PRT331	NUC453	500742815
38890	NUC163	PRT332	NUC453	500735594
38890	NUC164	PRT333	NUC453	500737569
38890	NUC165	PRT334	NUC453	500730242
38890	NUC166	PRT335	NUC453	500766374
38890	NUC167	PRT336	NUC453	500711885
40163			NUC743	244540

40975			NUC744	106061
40991			NUC745	106061
42896			NUC746	137110
43190			NUC747	244266
44053			NUC748	236316;249443;392450;449591; 486460;509697;519210;528872; 585226;589902;601550
45091			NUC749	238559
45179			NUC750	132269
45274	NUC84	PRT253	NUC404	1000867870
46679	NUC85	PRT254	NUC405	140265
47171			NUC751	420959
48024	NUC168	PRT337	NUC454	113448
48548			NUC752	227400
48603			NUC753	200687;244886
48670			NUC754	164887
48671			NUC755	221136
48823			NUC756	525888
48901			NUC757	525775
49018			NUC758	229481
49034			NUC759	237358
49133			NUC760	224706;582974
49140			NUC761	636146
49261			NUC762	186091
49387			NUC763	212526
49416			NUC764	530211
49426			NUC765	313150;313151
49493			NUC766	213393
49640			NUC767	626919
49863			NUC768	181361;382057;631692
49871			NUC769	181361;382057;631692
50015			NUC770	203070;331013
50049			NUC771	196104
50112			NUC772	118123;335719
50185			NUC773	489405;496185
50241			NUC774	145339;156186;244350
50353			NUC775	119282
50763			NUC776	500701668
50982			NUC777	637372
51130			NUC778	180402
51212			NUC779	238923
51346			NUC780	646118
51380	NUC169	PRT338	NUC455	523002
51400			NUC781	231102
51796			NUC782	210281
51954			NUC783	141971;183117;205855;502241
52076			NUC784	625028

TABLE II

<u>SEQ ID NO.</u> <u>in priority</u> <u>application</u>	<u>Chromosomal location</u>
51	3q26.2

<u>SEQ ID NO. in priority application</u>	<u>Chromosomal location</u>
452	2q31
483	11p11.2
505	20p12
573	5q21
796	15
1151	22q11.2
1590	X
2028	X
2932	8p23
3280	21
3326	15
3804	16p13.3
4172	1, 15q13-q14
4340	1
4609	4p13
4647	11q23, 11q23.3
4694	12p13.2
4733	14q32.1
4855	1q23-1q24
4908	12p13
5011	7q32, 7q32-36
5040	8p23
5089	4q11
5167	12
5278	19q13.2
5563	1, 2
5947	11q23
5974	4q28, 4q28-q31
6011	4
6290	2p13
6322	2p13
6345	16q24.3
6400	10q25-26
6892	4
7697	2p11, 4q28
8166	14q32
8666	16
8671	16
9755	3
10044	12q, 16
10322	1q11.1, 1q21
10584	3p21.3
10744	6
11019	11p11.2, 22q13
11342	7q21
12907	7q36.1
13202	15
13285	2q23-24, 11q23.2-24.2
27697	12p11.2
28517	2q31
28518	2q31
29673	16p13.3
29814	11

SEQ ID NO. in priority application	Chromosomal location
31422	11q13
31627	11
32962	9q32-33
33712	3p21.3, 3p21.31
35185	2
38631	15
48823	13q34-qter, 17
49018	7q22
49133	1q32
49387	17q21
50015	5q31
51380	Xq28

TABLE III

SEQ ID NO. in	Tissue Distribution
37	I:11
51	A:55 B:4 C:1 E:1 F:30 G:19 H:9
179	F:16 K:4
180	F:5 K:1
183	H:6
201	C:1
326	I:3
362	K:2
440	A:7
452	F:1 G:5
483	G:4
500	I:1
505	A:2
528	K:2
573	F:3 K:1
574	A:1
587	H:1 I:3
588	F:6 K:2
593	A:2 C:15 F:2 G:14
599	A:8 F:7 K:3
603	K:1
621	C:1
628	F:1 K:3
653	B:6 F:3 G:1 I:1 K:7
670	K:11
678	F:1 K:2
693	B:7 G:7
703	F:9 K:4
746	A:9 H:1 K:2
770	G:4
775	F:1 K:13
796	G:1
812	A:13 F:1 H:1
940	F:8
988	F:8 K:1
996	K:4

SEQ ID NO. in	Tissue Distribution
1036	A:13 I:11
1064	A:12 B:130 C:16 D:7 F:1 G:16
1151	A:9
1190	B:4 C:1 H:7 I:6
1458	J:55
1590	A:1
1853	A:1 G:1 H:1
1904	H:1
2028	A:8
2173	B:1 C:3 D:1 G:1
2368	K:4
2553	B:7 D:5 H:39
2556	B:7 D:5 H:39
2658	D:1
2690	K:2
2755	H:1
2800	F:8 K:6
2843	A:8 G:1 K:3
2852	F:1
2932	D:1
2955	K:1
3078	H:2
3280	F:4 K:3
3326	H:2
3387	H:7
3439	I:5
3501	A:10 C:1
3633	A:6 H:3
3678	K:2
3714	F:4 K:1
3796	F:1 K:1
3801	K:2
3804	B:1 C:10 D:1 G:2 I:11
3892	A:9
3985	B:1 C:5 D:1 H:1 J:6
4005	F:6 I:2 K:2
4063	A:2 B:10 C:4
4088	K:6
4111	A:11
4126	F:3 K:1
4172	F:3 K:2
4261	H:2
4340	A:3 G:3 I:1
4436	F:3
4609	D:10
4647	D:3
4660	D:1
4664	B:48 C:2 H:2 I:3
4678	C:5 D:17 G:4
4682	G:1
4687	D:5 F:2 H:1
4690	D:3 F:1 I:1 K:1
4694	I:3
4696	I:1
4733	D:1
4807	B:11 H:1
4809	I:1
4830	K:1

SEQ ID NO. in	Tissue Distribution
4855	H:1
4900	A:1
4908	A:16
4943	F:6 K:2
4947	K:3
4976	J:2
5000	A:2 B:14 C:17 D:5 F:1 G:9 H:3 I:3
5002	B:2
5005	C:31 F:1 J:2 K:2
5011	H:5 I:2
5040	B:1 C:2 D:11 G:1 J:1
5058	K:1
5071	F:1 K:1
5089	I:1
5117	H:7
5141	F:2 K:1
5162	A:1 B:8
5167	I:1
5178	G:1
5192	F:14 K:3
5214	A:10 B:19 C:5
5230	A:3 B:6 C:9 D:1 F:1 G:4 H:3 I:2
5240	A:1
5250	A:3 B:39 C:29 D:3 H:3 I:1 K:1
5262	A:3 B:39 C:34 D:3 H:3 I:1 K:1
5270	B:8
5278	D:1
5358	G:1
5453	A:2 C:6
5494	A:1 D:1
5499	A:3
5533	A:2 B:7 C:1
5563	D:1 F:1
5609	G:3
5657	G:1
5691	K:1
5748	A:3 B:1 C:4 D:1 G:30 H:1
5806	A:2 B:7 C:1 F:1 I:3 K:3
5816	B:5
5824	H:3
5861	B:16 C:27 D:13 F:7 H:21 I:12 K:4
5885	A:1
5913	K:1
5947	D:2
5966	G:11
5970	C:4
5974	D:12
5983	A:5
5985	B:2 C:5 H:1 I:1 K:2
6011	B:1
6080	B:1 I:1
6081	I:9
6108	F:5
6159	B:1
6231	A:1
6238	A:15 B:5 G:14 H:3 I:2
6252	G:7
6283	F:1 K:1

SEQ ID NO. in	Tissue Distribution
6290	B:3
6322	C:1
6329	A:1
6334	F:51 K:13
6345	A:9 G:1
6350	A:1 B:2
6358	B:9 H:1
6384	B:1 C:1
6400	K:1
6418	K:1
6431	A:1 B:1 K:1
6453	F:1
6636	A:6 B:37 C:11 D:1 F:1 G:4 H:1 J:1
6660	F:1 K:2
6688	A:1 B:42 D:1 G:3 I:3 K:3
6727	F:16 G:5 K:103
6835	K:1
6865	A:1 I:1
6892	B:23 C:1 H:2 I:1 J:12
7000	B:1
7041	C:1 F:17 H:1
7533	G:2
7535	A:5
7577	F:1
7697	D:10 F:3
7712	F:1 K:3
8009	B:2 C:2 H:1
8078	F:5 K:20
8097	I:3
8166	A:7 B:6 D:8 G:52
8262	H:1
8341	A:1 F:2 G:3 H:2 I:1
8534	F:1 K:6
8666	B:4 D:1 G:16 K:1
8671	G:1 K:2
8744	H:1
8968	A:12 G:3 H:1
8994	F:1 K:4
9297	D:1
9327	K:1
9332	K:4
9406	B:10 D:3 H:39
9407	B:10 D:3 H:39
9668	F:9 K:41
9679	B:1 F:1 K:1
9755	K:5
9868	F:9 K:1
10044	A:10 B:10 D:14 F:2 G:24 H:5 K:4
10322	B:1 F:2 K:3
10526	B:3 H:3
10584	A:2
10650	K:3
10739	A:1 B:1 G:3 J:1
10743	C:1 J:1
10744	A:6 B:3 G:2 I:3
10761	H:1
10880	D:1 G:1
10942	A:30

SEQ ID NO. in	Tissue Distribution
11019	F:4
11278	I:1
11342	F:1 K:2
11562	F:1 K:4
11688	F:2 K:12
11735	A:25 B:57 C:7 D:1 F:1 G:1 K:1
11813	A:8
12039	B:3
12043	A:6
12048	F:1 H:1
12098	A:4 B:8 H:2 I:1
12202	K:2
12220	C:2
12243	A:1 B:132 C:38 D:31 F:6 G:11
12263	H:4
12276	A:18
12490	A:9 B:11 C:8 G:4
12604	A:6 B:14 C:22 D:1 F:1 G:4 H:4 I:1
12657	B:1 C:6 J:2
12788	K:2
12901	G:2
12907	I:6
13013	D:5
13202	A:2 B:14 C:9 D:1 G:3 H:3 J:3 K:1
13229	A:3
13256	F:2 K:5
13267	I:8
13285	B:1 C:42 D:5 I:1 J:10
26638	D:2
26710	A:90 B:48 G:27
26726	C:4
26786	A:24
26982	A:1 G:1
27084	F:16 K:83
27273	C:3
27301	A:3 B:3 H:2
27336	B:1 G:1
27361	F:2
27374	B:1 D:1
27627	F:1 K:1
27697	A:1 C:1
27877	H:1 I:19
28413	H:1
28517	K:2
28518	G:3
29120	I:1
29469	A:2 G:2
29472	A:1
29557	C:1
29673	D:1
29814	A:1
30218	F:7 K:1
30446	B:1 C:1 H:10
30477	G:7
30583	K:5
30719	A:1 B:9 I:3
31356	H:2
31422	G:1

SEQ ID NO. in	Tissue Distribution
31554	K:1
31627	C:1
31726	C:1
31744	F:1
31790	A:25
32102	G:1
32473	B:1
32475	B:1
32962	G:1
33130	K:3
33712	B:1 C:1
35005	A:1
35185	C:1
35258	K:1
35326	B:1 F:3 G:1 H:1 I:1 K:3
35597	F:1 K:1
35912	A:1
35984	F:1
36122	K:1
37337	F:1 K:2
38112	K:1
38220	A:1
38311	A:1
38631	K:1
38749	H:1
38890	G:21
40163	K:2
40975	B:8 D:1 H:39
40991	B:8 D:1 H:39
42896	H:1
43190	K:1
44053	C:11 D:1 H:1 I:2 J:5
45091	J:1
45179	H:1
45274	C:4 H:2
46679	D:3 F:1 H:1
47171	C:1
48024	B:4 C:5
48548	F:1
48603	F:41 K:19
48670	K:1
48671	K:2
48823	C:65
48901	B:1 G:2 I:2
49018	B:2 C:2 F:2 J:2
49034	D:1 F:2 H:1 I:3 K:2
49133	F:4 K:2
49140	A:1 C:1
49261	F:3 K:23
49387	F:2 K:1
49416	B:18 C:1
49426	J:51
49493	C:1 F:1 I:6
49640	A:5 F:1
49863	A:6 C:2 I:1
49871	A:6 C:2 I:1
50015	K:2
50049	K:2

SEQ ID NO. in	Tissue Distribution
50112	K:3
50185	I:6
50241	F:3 H:1
50353	F:1 K:1
50763	A:14
50982	A:2
51130	K:2
51212	J:2
51346	A:2 B:1 D:1
51380	B:1 F:1
51400	F:3
51796	F:1 K:2
51954	B:13 C:1 D:1 F:2 K:1
52076	A:2

TABLE IV

SEQ ID NO. in priority application	Tissue source
37	adenocarcinoma(2), carcinoid(2), testis(1), tonsil(1)
51	adipose tissue, white(4), cerebellum(3), cochlea(1), colon tumor rer+(2), dorsal root ganglion(1), hippocampus(1), kidney(1), liver(1), malignant melanoma, metastatic to lymph node(1), muscle(2), normal leg muscle(1), parathyroid tumor(1), pectoral muscle (after mastectomy)(1), placenta(1), substantia nigra(1), total brain(1)
201	pancreas(1)
326	ovarian tumor(1), uterus(1)
440	brain cortex(1), carcinoid tumor(1)
483	pbl(1), adenocarcinoma(1), astrocytoma(1), ovarian tumor(1), schizophrenic brain s-11 frontal lobe(1)
500	colon(4), colon tumor rer+(1), pooled germ cell tumors(1)
505	total brain(1)
528	brain(2), cerebellum(1), colon(1), ovarian tumor(6)
573	melanoma (mewo cell line)(1)
587	germinal center b cell(1), lymphoma(1), parathyroid tumor(1)
593	ovarian tumor(1), placenta(1)
621	placenta.(2)
653	2 pooled tumors (clear cell type)(2), anaplastic oligodendroglioma(2), glioblastoma (pooled)(2)
678	ovarian tumor(1), prostate(1)
693	brain(1), placenta(1)
703	small cell carcinoma(2)
746	small cell carcinoma(1)
770	colon(1), frontal lobe(1), human pancreatic islets(1), normal leg muscle(1), ovarian tumor(1), pancreatic islet(2), senescent fibroblast(1), total brain(1)
775	anaplastic oligodendroglioma(1), frontal lobe(1)
796	adrenal adenoma(1), adrenal gland(1), breast tumor(3), placenta(2)
812	heart(1), brain(1), frontal lobe(3), neuroepithelial cells(1), retina(1), small cell carcinoma(1), total brain(1)
988	germinal center b cell(1)
1064	2 pooled tumors (clear cell type)(1), ewing's sarcoma(2), adenocarcinoma(5), anaplastic oligodendroglioma(5), breast tumor(1), carcinoid(4), colon(3), colon tumor(1), colon tumor

	rer+(2), frontal lobe(4), germinal center b cell(7), kidney tumor(1), lung tumor(1), metastatic prostate bone lesion(2), ovarian tumor(6), parathyroid tumor(7), pectoral muscle (after mastectomy)(12), placenta(1), pooled germ cell tumors(4), senescent fibroblast(2), squamous cell carcinoma from base of tongue(1), tumor, 5 pooled (see description)(1)
1151	anaplastic oligodendroglioma(4), carcinoid(1), colon tumor rer+(1), medulloblastoma(1), normal prostate(1), parathyroid tumor(1), tumor(1)
1190	colon(1)
1853	2 pooled high-grade transitional cell tumors(1), 2 pooled tumors (clear cell type)(1)
2173	2 pooled tumors (clear cell type)(2), b-cell, chronic lymphocytic leukemia(2), cd34+, cd38- from normal bone marrow donor(6), adenocarcinoma(1), alveolar rhabdomyosarcoma(1), anaplastic oligodendroglioma(3), breast(1), carcinoid(1), cerebellum(2), cochlea(3), colon(11), colon tumor rer+(2), dorsal root ganglion(1), early stage papillary serous carcinoma(1), epithelium (cell line)(1), follicular lymphoma(1), frontal lobe(15), germinal center b-cells(3), invasive adenocarcinoma(5), kidney(1), kidney tumor(2), larynx(1), liver(1), lung carcinoma(1), lymphoma(1), meningioma(1), moderately differentiated adenocarcinoma(2), moderately-differentiated adenocarcinoma(1), muscle(1), normal prostate(6), normal prostatic epithelial cells(1), oligodendroglioma(3), ovarian tumor(4), papillary serous carcinoma(1), pectoral muscle (after mastectomy)(23), pooled germ cell tumors(2), prostate(1), stem cell 34+/38+(2), thyroid(1), tumor, 5 pooled (see description)(4), two pooled squamous cell carcinomas(2)
2553	brain(2), tumor(1)
2556	brain(2), tumor(1)
2755	frontal lobe(1)
2843	b-cell, chronic lymphocytic leukemia(2), anaplastic oligodendroglioma(5), carcinoid(1), placenta(1)
2852	frontal lobe(3)
2932	bone marrow(14), brain(1), hematopoietic from aml patient(1), liver(2), normal cortical stroma(1)
3078	frontal lobe(2)
3280	b-cell, chronic lymphocytic leukemia(1), germinal center b cell(1), leiomyosarcoma(1), pooled germ cell tumors(1)
3326	muscle(3), normal leg muscle(1), parathyroid tumor(1)
3387	carcinoid(1), pooled germ cell tumors(4)
3439	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(3), ewing's sarcoma(1), anaplastic oligodendroglioma(1), colon(2), glioblastoma (pooled)(1), mantle cell lymphoma(1), parathyroid tumor(1), senescent fibroblast(1), squamous cell carcinoma(1)
3501	anaplastic oligodendroglioma(1), breast(1), carcinoid(2), colon(1), colon tumor rer+(1), epithelium (cell line)(1), kidney tumor(1), ovarian tumor(1), parathyroid tumor(2), pooled germ cell tumors(4), senescent fibroblast(1), squamous cell carcinoma(1), testis(1), tumor(2)
3633	cerebellum(1), muscle(1), retina(4), small cell carcinoma(1), total brain(3)
3714	2 pooled tumors (clear cell type)(1), colon(1)
3804	heart(1), anaplastic oligodendroglioma(1), carcinoid(1), colon tumor(2), germinal center b cell(2), kidney tumor(1), lung tumor(3), lymphoid(2), moderately differentiated adenocarcinoma(1), muscle(1), ovarian tumor(3), pectoral muscle (after mastectomy)(1), squamous cell carcinoma(3)
3892	blood(1), total brain(2)
3985	anaplastic oligodendroglioma(1), breast(1)
4005	2 pooled tumors (clear cell type)(1), brain(2), germinal center b cell(1), muscle(1)
4063	cerebral cortex(1), brain(1), carcinoid(2), cerebellum(1), schizophrenic brain s-11 frontal lobe(1), senescent fibroblast(1), total brain(2)
4111	anaplastic oligodendroglioma(1), cerebellum(1), colon(1)
4172	retina(3)
4340	heart(1), blood(1), brain(2), colon(6), frontal lobe(9), invasive tumor (cell line)(1), melanocyte(1), neuroepithelial cells(1), senescent fibroblast(1), small cell carcinoma(2), total brain(3)
4436	b-cell, chronic lymphocytic leukemia(1), bulk tumor(1), colon(24), early stage papillary serous carcinoma(3), lung carcinoma(1), pancreatic cancer(1), pooled germ cell tumors(1)
4609	blood(1)
4647	colon(1), liver(1)

4660	adenocarcinoma(1)
4664	2 pooled tumors (clear cell type)(5), adenocarcinoma(3), brain(1), breast(1), colon(4), colon tumor rer+(1), frontal lobe(5), liver(1), neuroepithelial cells(1), normal prostate(1), ovarian tumor(1), ovary(1), total brain(1), tumor(1)
4678	colon tumor, rer+(2)
4682	2 pooled tumors (clear cell type)(4), anaplastic oligodendroglioma(2), breast(3), carcinoid(1), glioblastoma (pooled)(1), pooled germ cell tumors(1)
4687	2 pooled tumors (clear cell type)(1), carcinoid(3), colon(3), normal prostate(1), pooled germ cell tumors(1)
4690	carcinoid(1), colon(3), normal prostate(1), pooled germ cell tumors(1)
4694	colon tumor, rer+(1)
4733	colon(3), liver(14), pancreatic islet(1)
4807	2 pooled tumors (clear cell type)(1), alveolar rhabdomyosarcoma(1), carcinoid(3), colon(2), normal prostate(2), normal prostatic epithelial cells(1), ovarian tumor(23), prostate(2), serous adenocarcinoma(1), total brain(2), tumor(1), tumor, 5 pooled (see description)(4)
4809	2 pooled tumors (clear cell type)(1), alveolar rhabdomyosarcoma(1), carcinoid(3), colon(2), normal prostate(2), normal prostatic epithelial cells(1), ovarian tumor(28), prostate(2), serous adenocarcinoma(1), total brain(2), tumor(1), tumor, 5 pooled (see description)(5)
4855	b-cell, chronic lymphocytic leukemia(1), carcinoid(1)
4900	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(1), colon(1), liver(1), low-grade prostatic neoplasia(1), normal prostate(1), pooled germ cell tumors(1)
4908	alveolar rhabdomyosarcoma(1), colon(1), hemopoietic system(1), liver(1), malignant ascitic effusion(1), moderately-differentiated adenocarcinoma(1), neuroepithelial cells(1), parathyroid tumor(1), synovial membrane(1), uterus(1)
4947	parathyroid tumor(1), testis(2)
5000	2 pooled tumors (clear cell type)(3), b-cell, chronic lymphocytic leukemia(2), anaplastic oligodendroglioma(5), breast(4), carcinoid(2), colon tumor, rer+(1), germ cell tumor(1), germinal center b cell(6), invasive prostate tumor(1), lung tumor(1), metastatic prostate bone lesion(2), normal prostate(1), parathyroid tumor(2), pectoral muscle (after mastectomy)(3), pooled germ cell tumors(1), senescent fibroblast(4), stroma(1), thyroid(1), tumor, 5 pooled (see description)(2)
5002	pectoral muscle (after mastectomy)(1), senescent fibroblast(1)
5005	anaplastic oligodendroglioma(1), germinal center b cell(1), lobular carcinoma in situ(1), lymphoma(1), oligodendroglioma(1), pooled germ cell tumors(8)
5011	breast cancer(1), normal prostate(5), seminal vesicles(1)
5040	bone marrow(18), brain(1), hematopoietic from aml patient(1), liver(2), normal cortical stroma(1)
5089	parotid gland(1)
5117	adenocarcinoma(2), breast(2), carcinoid(1), colon(3), colon tumor rer+(1), epithelium (cell line)(3), moderately differentiated adenocarcinoma(1), moderately-differentiated adenocarcinoma(2), normal prostate(3), normal prostatic epithelial cells(2), placenta(1), prostate(3), small cell carcinoma(1), squamous cell carcinoma(2), squamous cell carcinoma from base of tongue(1)
5162	hippocampus(1), schizophrenic brain s-11 frontal lobe(1)
5167	colon(3), colon tumor rer+(2), pooled germ cell tumors(1)
5214	kidney(1), ovarian tumor(1), pituitary gland(1), total brain(2)
5230	2 pooled tumors (clear cell type)(5), placenta(1), breast(1), breast tumor(1), carcinoid(2), colon(1), colon carcinoma(1), colon mucosa(1), colon tumor rer+(4), germinal center b cell(1), kidney tumor(1), normal prostate(5), papillary serous carcinoma(1), parathyroid tumor(3), pectoral muscle (after mastectomy)(3), pooled germ cell tumors(2), senescent fibroblast(1), squamous cell carcinoma from base of tongue(1)
5240	bone(1), colon(2), frontal lobe(2), glioblastoma (pooled)(1), liver(1), ovarian tumor(1), parathyroid tumor(1), tumor, 5 pooled (see description)(1)
5250	2 pooled tumors (clear cell type)(1), anaplastic oligodendroglioma(2), brain(5), carcinoid(1), colon(2), colon tumor rer+(1), pooled germ cell tumors(1)
5262	2 pooled tumors (clear cell type)(1), anaplastic oligodendroglioma(2), brain(5), carcinoid(1), colon(2), colon tumor rer+(1), pooled germ cell tumors(1)
5270	ovarian tumor(1)
5278	2 pooled tumors (clear cell type)(7), anaplastic oligodendroglioma(7), breast(1), breast

	tumor(4), carcinoid(3), colon(1), colon tumor rer+(2), glioblastoma (pooled)(1), liver(3), pooled germ cell tumors(6), thyroid(2)
5358	2 pooled tumors (clear cell type)(1), anaplastic oligodendroglioma(2), brain(2), colon(2), germinal center b cell(4), melanocyte(1), normal prostate(3), parathyroid tumor(2), pectoral muscle (after mastectomy)(3), pooled germ cell tumors(4), senescent fibroblast(2)
5494	2 pooled tumors (clear cell type)(2), cd34+, cd38- from normal bone marrow donor(2), anaplastic oligodendroglioma(2), brain(2), colon(2), colon tumor rer+(1), early stage papillary serous carcinoma(1), germinal center b cell(3), glioblastoma (pooled)(3), moderately-differentiated adenocarcinoma(1), normal prostate(1), normal prostatic epithelial cells(1), omentum(1), ovarian tumor(4), ovary(1), parathyroid tumor(3), pectoral muscle (after mastectomy)(1), pooled germ cell tumors(1), senescent fibroblast(2), stem cell 34+/38+(1), synovial membrane(1), synovial sarcoma(1), tumor(1)
5499	2 pooled tumors (clear cell type)(1), brain(2), pancreatic islet(2)
5533	2 pooled tumors (clear cell type)(2), anaplastic oligodendroglioma(3), breast(1), carcinoid(1), germinal center b cell(2), glioblastoma (pooled)(2), pooled germ cell tumors(2), senescent fibroblast(1), small cell carcinoma(2)
5563	colon(3), kidney(1), liver(1), neuroepithelial cells(1), normal prostatic epithelial cells(1), ovarian tumor(1), senescent fibroblast(1), total brain(1)
5691	2 pooled high-grade transitional cell tumors(1), 2 pooled tumors (clear cell type)(3), adenocarcinoma(3), anaplastic oligodendroglioma(2), carcinoid(3), colon(2), germinal center b cell(1), glioblastoma (pooled)(3), medulloblastoma(1), ovarian tumor(1), parathyroid tumor(2), pectoral muscle (after mastectomy)(2), prostate(1), three pooled meningiomas(1), total brain(1)
5748	bone(2), anaplastic oligodendroglioma(5), breast(1), carcinoid(1), colon tumor rer+(1), frontal lobe(2), germinal center b cell(1), glioblastoma (pooled)(1), ovarian tumor(4), parathyroid tumor(2), pooled germ cell tumors(1), senescent fibroblast(1), tumor, 5 pooled (see description)(1)
5806	2 pooled tumors (clear cell type)(4), female, 19 years old, normal leg muscle(2), alveolar rhabdomyosarcoma(1), anaplastic oligodendroglioma(4), breast(1), carcinoid(3), germinal center b cell(1), glioblastoma (pooled)(2), normal prostate(1)
5816	anaplastic oligodendroglioma(1), frontal lobe(4)
5824	2 pooled tumors (clear cell type)(2), b-cell, chronic lymphocytic leukemia(2), adenocarcinoma(1), anaplastic oligodendroglioma(3), blood(1), breast(1), carcinoid(9), cerebellum(1), colon(3), fibrotheoma(1), follicular lymphoma(1), germinal center b cell(2), glioblastoma (pooled)(2), kidney tumor(1), low-grade prostatic neoplasia(2), normal prostatic epithelial cells(1), ovarian tumor(2), parathyroid tumor(7), pooled germ cell tumors(1), senescent fibroblast(6), thyroid(1)
5861	adipose tissue, white(2), bone marrow from femur(1), carcinoid(1), epithelium(1), normal prostatic epithelial cells(1), pectoral muscle (after mastectomy)(1)
5885	2 pooled tumors (clear cell type)(1), female, 19 years old, normal leg muscle(1), anaplastic oligodendroglioma(5), bone marrow stroma(1), colon tumor rer+(1), germinal center b cell(5), glioblastoma (pooled)(1), kidney tumor(1), melanocyte(2), moderately differentiated adenocarcinoma(1), moderately-differentiated adenocarcinoma(1), normal prostate(1), normal prostatic epithelial cells(2), ovarian tumor(2), parathyroid tumor(1), senescent fibroblast(1), three pooled meningiomas(1), tumor(1)
5947	2 pooled tumors (clear cell type)(4), adenocarcinoma(1), colon tumor(3), intestine(1), liver(1)
5970	b-cell, chronic lymphocytic leukemia(1), carcinoid(1), germinal center b cell(2), pooled germ cell tumors(3)
5974	blood(5), bone marrow(1), liver(4), reticulocyte(1)
5983	2 pooled tumors (clear cell type)(1), adenocarcinoma(1), carcinoid(1), germinal center b cell(4), medulloblastoma(1), pooled germ cell tumors(1), tumor, 5 pooled (see description)(1)
5985	2 pooled tumors (clear cell type)(3), b-cell, chronic lymphocytic leukemia(2), anaplastic oligodendroglioma(5), breast(4), carcinoid(2), colon tumor, rer+(1), germ cell tumor(1), germinal center b cell(6), invasive prostate tumor(1), lung tumor(1), metastatic prostate bone lesion(2), normal prostate(1), parathyroid tumor(2), pectoral muscle (after mastectomy)(3), pooled germ cell tumors(1), senescent fibroblast(4), stroma(1), thyroid(1), tumor, 5 pooled (see description)(2)
6011	brain(5), carcinoid(1), frontal lobe(1), lung carcinoma(1), retina(1)
6080	2 pooled tumors (clear cell type)(8), ewing's sarcoma(2), heart(1), adenocarcinoma(1), alveolar rhabdomyosarcoma(6), anaplastic oligodendroglioma(5), aorta(1), breast(4), bulk

	germ cell seminoma(2), colon(3), germinal center b cell(2), glioblastoma (pooled)(1), kidney(2), lung carcinoma(1), metastatic prostate bone lesion(4), normal prostate(1), normal prostatic epithelial cells(1), oligodendroglioma(1), ovary(2), parathyroid tumor(5), pectoral muscle (after mastectomy)(19), pooled germ cell tumors(1), prostate(1), senescent fibroblast(4), tumor, 5 pooled (see description)(1)
6081	brain(1), germinal center b cell(1), testis(1)
6108	2 pooled tumors (clear cell type)(1), carcinoid(1), germinal center b cell(1), parathyroid tumor(1)
6159	2 pooled tumors (clear cell type)(2), ewing's sarcoma(1), schwannoma tumor(1), adipose tissue, white(1), adrenal adenoma(3), amygdala(1), anaplastic oligodendroglioma(1), astrocytoma(2), bone marrow stroma(3), borderline ovarian carcinoma(1), cochlea(3), colon tumor(1), epithelium (cell line)(4), frontal lobe(3), germinal center b cell(1), human pancreatic islets(1), kidney tumor(3), larynx(1), liver(1), lung carcinoma(1), lung tumor(1), normal leg muscle(1), oligodendroglioma(1), ovarian tumor(8), parathyroid tumor(1), pectoral muscle (after mastectomy)(8), prostate tumor(1), senescent fibroblast(3), small cell carcinoma(4)
6231	adenocarcinoma(2), breast(3), colon(4), colon tumor(2), endometrioid ovarian metastasis(1), epithelium (cell line)(22), frontal lobe(1), germ cell tumor(3), invasive tumor (cell line)(13), moderately differentiated adenocarcinoma(1), normal prostatic epithelial cells(1), ovarian tumor(4), ovary(1), pancreatic islet(1), placenta(trophoblast)(1), pooled germ cell tumors(2), squamous cell carcinoma(1), synovial sarcoma(1), tumor(3), tumor, 5 pooled (see description)(2), two pooled squamous cell carcinomas(2)
6238	bone(1), colon(2), ovarian tumor(2), small cell carcinoma(2), total brain(1)
6252	2 pooled tumors (clear cell type)(1), ewing's sarcoma(6), adenocarcinoma(1), alveolar rhabdomyosarcoma(3), anaplastic oligodendroglioma(2), brain(1), breast(2), breast tumor(1), carcinoid(2), cochlea(2), germ cell tumor(1), glioblastoma (pooled)(1), kidney(1), kidney tumor(6), liposarcoma(1), lung carcinoma(1), lymphoma(1), metastatic prostate bone lesion(6), muscle(1), normal prostate(1), normal prostatic epithelial cells(1), ovarian tumor(2), ovary(4), parathyroid tumor(1), pectoral muscle (after mastectomy)(5), pooled germ cell tumors(4), prostate(1), renal cell tumor(1), senescent fibroblast(3), stem cells(1), tumor, 5 pooled (see description)(4)
6290	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(1), heart(3), lymphoma(1), adenocarcinoma(1), adipose tissue, white(1), brain(3), carcinoid(5), cerebellum(5), colon(8), epithelium (cell line)(1), frontal lobe(1), germinal center b-cells(1), kidney tumor(1), medulloblastoma(1), melanoma (mewo cell line)(1), normal prostate(1), omentum(1), ovarian tumor(1), pancreatic islet(1), placenta(3), pooled frontal lobe(1), retina(1), retinal fovea(1), schizophrenic brain s-11 frontal lobe(1), senescent fibroblast(1), small cell carcinoma(3), synovial membrane(1), total brain(2)
6322	b-cell, chronic lymphocytic leukemia(1), heart(4), lymphoma(1), adipose tissue, white(1), brain(3), carcinoid(4), cerebellum(5), colon(8), epithelium (cell line)(1), frontal lobe(1), germinal center b-cells(1), kidney tumor(1), melanoma (mewo cell line)(1), omentum(1), ovarian tumor(1), pancreatic islet(1), placenta(3), pooled frontal lobe(1), retinal fovea(1), schizophrenic brain s-11 frontal lobe(1), senescent fibroblast(1), small cell carcinoma(3), synovial membrane(1), total brain(2)
6329	b-cell, chronic lymphocytic leukemia(3), heart(3), lymphoma(2), adenocarcinoma(6), adipose tissue, white(1), adrenal adenoma(2), anaplastic oligodendroglioma(2), bone marrow stroma(1), brain(2), breast(1), carcinoid(7), cerebellum(5), colon(8), epithelium (cell line)(1), frontal lobe(1), germ cell tumor(2), germinal center b-cells(1), kidney tumor(1), larynx(1), lung tumor(1), medulloblastoma(1), melanoma (mewo cell line)(1), metastatic melanoma to bowel(1), moderately differentiated adenocarcinoma(1), normal prostate(2), omentum(1), ovarian tumor(1), pancreatic islet(2), papillary serous ovarian metastasis(1), parathyroid tumor(1), pectoral muscle (after mastectomy)(1), placenta(2), pooled frontal lobe(1), pooled germ cell tumors(3), prostate(1), retinal fovea(1), schizophrenic brain s-11 frontal lobe(1), senescent fibroblast(3), small cell carcinoma(5), squamous cell carcinoma(1), synovial membrane(1), total brain(2), tumor(1), tumor, 5 pooled (see description)(1), two pooled squamous cell carcinomas(1)
6334	2 pooled tumors (clear cell type)(3), b-cell, chronic lymphocytic leukemia(2), anaplastic oligodendroglioma(9), glioblastoma (pooled)(3), kidney(1), normal prostate(3), ovarian tumor(1), senescent fibroblast(2)
6345	testis(3)

6350	2 pooled tumors (clear cell type)(2), b-cell, chronic lymphocytic leukemia(1), bone(1), adenocarcinoma(1), anaplastic oligodendroglioma(10), breast(1), breast tumor(1), colon(2), glioblastoma (pooled)(6), kidney(1), lung carcinoma(1), medulloblastoma(1), metastatic prostate bone lesion(1), normal prostate(1), pectoral muscle (after mastectomy)(4), pooled germ cell tumors(2), senescent fibroblast(2), squamous cell carcinoma(1), testis(1), tumor(1), tumor, 5 pooled (see description)(1)
6358	2 pooled tumors (clear cell type)(2), anaplastic oligodendroglioma(4), brain(4), carcinoid(2), colon(3), colon tumor rer+(1), germinal center b cell(4), glioblastoma (pooled)(2), normal prostate(1), normal prostatic epithelial cells(1), parathyroid tumor(2), pectoral muscle (after mastectomy)(1), senescent fibroblast(3)
6384	epithelium(1), meningioma(1), parathyroid tumor(2), senescent fibroblast(1)
6400	blood(2), brain(2), colon(6), parathyroid tumor(1), tumor(1)
6431	adenocarcinoma(1)
6453	pooled germ cell tumors(3)
6636	2 pooled tumors (clear cell type)(1), ewing's sarcoma(2), adenocarcinoma(3), anaplastic oligodendroglioma(3), breast tumor(1), carcinoid(4), colon(2), colon tumor rer+(2), frontal lobe(3), germinal center b cell(6), kidney tumor(1), ovarian tumor(4), parathyroid tumor(4), pectoral muscle (after mastectomy)(10), placenta(1), pooled germ cell tumors(2), senescent fibroblast(1), squamous cell carcinoma from base of tongue(1), tumor, 5 pooled (see description)(1)
6688	b-cell, chronic lymphocytic leukemia(3), alveolar rhabdomyosarcoma(1), anaplastic oligodendroglioma(1), breast(2), carcinoid(1), colon(1), colon tumor(1), four pooled pituitary adenomas(1), germinal center b cell(2), glioblastoma (pooled)(1), kidney tumor(2), moderately-differentiated adenocarcinoma(1), muscle(2), pectoral muscle (after mastectomy)(1), pooled germ cell tumors(3), senescent fibroblast(3), synovial sarcoma(1), testis(1), tumor, 5 pooled (see description)(3)
6727	frontal lobe(1), schizophrenic brain s-11 frontal lobe(1)
6865	frontal lobe(4), germinal center b cell(1), muscle(1), ovarian tumor(2), pectoral muscle (after mastectomy)(3), senescent fibroblast(1), small cell carcinoma(1), thyroid(1)
6892	2 pooled tumors (clear cell type)(5), adenocarcinoma(1), anaplastic oligodendroglioma(5), brain(5), breast(3), breast tumor(1), carcinoid(5), cerebellum(1), colon(3), colon tumor rer+(2), frontal lobe(5), germinal center b cell(3), glioblastoma (pooled)(2), moderately-differentiated adenocarcinoma(1), normal prostate(3), ovary(2), parathyroid tumor(3), pectoral muscle (after mastectomy)(1), placenta(1), pooled germ cell tumors(5), senescent fibroblast(3), tumor(1), tumor, 5 pooled (see description)(1)
7000	anaplastic oligodendroglioma(2), frontal lobe(2)
7041	germinal center b cell(1), total brain(2)
7533	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(1), schwannoma tumor(1), anaplastic oligodendroglioma(1), astrocytoma(1), breast(1), cochlea(1), colon tumor rer+(2), germ cell tumor(1), germinal center b cell(2), glioblastoma (pooled)(3), hepatoma(4), medulloblastoma(1), metastatic prostate bone lesion(1), moderately-differentiated adenocarcinoma(1), ovarian tumor(1), prostate tumor(1), senescent fibroblast(1), squamous cell carcinoma(1), three pooled meningiomas(1)
7535	b-cell, chronic lymphocytic leukemia(1), schwannoma tumor(1), anaplastic oligodendroglioma(1), astrocytoma(1), breast(1), cochlea(1), colon tumor rer+(2), germ cell tumor(1), germinal center b cell(2), glioblastoma (pooled)(3), hepatoma(4), medulloblastoma(1), metastatic prostate bone lesion(1), moderately-differentiated adenocarcinoma(1), ovarian tumor(1), senescent fibroblast(1), squamous cell carcinoma(1)
7697	colon(1), invasive adenocarcinoma(3), liver(1), moderately-differentiated adenocarcinoma(1)
8009	alveolar rhabdomyosarcoma(2), anaplastic oligodendroglioma(1), carcinoid(9), colon(1), colon tumor rer+(1), germinal center b cell(2), glioblastoma (pooled)(1), normal prostate(1), ovary(2), pooled germ cell tumors(2), thyroid(2), tumor(1)
8078	frontal lobe(2)
8079	b-cell, chronic lymphocytic leukemia(2), bone(1), anaplastic oligodendroglioma(1), colon(1), pooled germ cell tumors(3)
8097	colon(1)
8166	2 pooled tumors (clear cell type)(5), adenocarcinoma(6), anaplastic oligodendroglioma(3), breast tumor(1), carcinoid(1), colon(1), epithelium (cell line)(1), glioblastoma (pooled)(2), lung tumor(2), metastatic melanoma to bowel(1), moderately-differentiated adenocarcinoma(1), normal prostate(1), ovary(1), papillary serous ovarian metastasis(1),

	parathyroid tumor(1), pooled germ cell tumors(4), renal cell tumor(1), squamous cell carcinoma(5), synovial sarcoma(1), tumor(3), tumor, 5 pooled (see description)(3)
8341	2 pooled tumors (clear cell type)(1), anaplastic oligodendroglioma(2), germinal center b cell(3), glioblastoma (pooled)(1), normal prostate(1), oligodendroglioma(1), pectoral muscle (after mastectomy)(2), tumor(1)
8534	germinal center b cell(4)
8666	2 pooled high-grade transitional cell tumors(1), 2 pooled tumors (clear cell type)(5), b-cell, chronic lymphocytic leukemia(2), adenocarcinoma(3), alveolar rhabdomyosarcoma(3), amygdala(1), anaplastic oligodendroglioma(3), blood(2), breast(1), carcinoid(3), colon(3), germinal center b cell(4), muscle(1), normal prostate(1), ovarian tumor(6), parathyroid tumor(2), pectoral muscle (after mastectomy)(4), pheochromocytoma(1), placenta(1), senescent fibroblast(11), two pooled squamous cell carcinomas(1)
8671	2 pooled high-grade transitional cell tumors(1), 2 pooled tumors (clear cell type)(3), b-cell, chronic lymphocytic leukemia(1), adenocarcinoma(6), alveolar rhabdomyosarcoma(3), amygdala(2), anaplastic oligodendroglioma(2), blood(2), breast(2), carcinoid(1), colon(2), germinal center b cell(5), liver cancer(1), metastatic prostate bone lesion(1), muscle(1), normal prostate(1), ovarian tumor(5), parathyroid tumor(2), pectoral muscle (after mastectomy)(5), placenta(2), pooled germ cell tumors(1), renal cell tumor(1), senescent fibroblast(8)
8968	b-cell, chronic lymphocytic leukemia(1), neuroepithelial cells(1), total brain(1)
9406	brain(2), colon(1), tumor(1)
9407	brain(2), colon(1), tumor(1)
9668	2 pooled tumors (clear cell type)(8), adenocarcinoma(1), anaplastic oligodendroglioma(1), breast carcinoma in situ(3), cerebellum(1), oligodendroglioma(1), papillary serous carcinoma(2), parathyroid tumor(26), pooled germ cell tumors(1)
9679	ovarian tumor(3), parathyroid tumor(1), uterus(1)
9755	2 pooled tumors (clear cell type)(1), adenocarcinoma(1), carcinoid(1), ovarian tumor(1), pooled germ cell tumors(3), senescent fibroblast(2), total brain(2)
9868	brain(1), senescent fibroblast(1)
10044	2 pooled high-grade transitional cell tumors(1), 2 pooled tumors (clear cell type)(3), b-cell, chronic lymphocytic leukemia(1), adenocarcinoma(6), alveolar rhabdomyosarcoma(3), amygdala(2), anaplastic oligodendroglioma(2), blood(1), breast(2), carcinoid(1), colon(2), germinal center b cell(5), liver cancer(1), metastatic prostate bone lesion(1), normal prostate(1), osteosarcoma(1), ovarian tumor(5), parathyroid tumor(2), pectoral muscle (after mastectomy)(5), placenta(2), pooled germ cell tumors(1), renal cell tumor(1), senescent fibroblast(8)
10322	2 pooled tumors (clear cell type)(9), b-cell, chronic lymphocytic leukemia(2), anaplastic oligodendroglioma(4), breast(1), carcinoid(3), colon(3), colon tumor rer+(1), germinal center b cell(1), glioblastoma (pooled)(4), kidney tumor(1), low-grade prostatic neoplasia(1), metastatic melanoma to bowel(1), normal prostate(2), ovary bulk tumor(1), parathyroid tumor(3), pectoral muscle (after mastectomy)(3), pooled germ cell tumors(2), senescent fibroblast(2), small cell carcinoma(2), synovial sarcoma(2), two pooled squamous cell carcinomas(3)
10526	2 pooled tumors (clear cell type)(2), ewing's sarcoma(4), adenocarcinoma(1), adipose tissue, white(1), alveolar rhabdomyosarcoma(3), anaplastic oligodendroglioma(5), brain(1), breast(4), breast tumor(1), carcinoid(2), colon(10), epithelium (cell line)(1), frontal lobe(7), germ cell tumor(1), glioblastoma (pooled)(1), invasive prostate tumor(1), kidney tumor(8), lymphoma(4), metastatic prostate bone lesion(2), muscle(3), normal prostate(2), normal prostatic epithelial cells(3), ovarian tumor(1), ovary(2), parathyroid tumor(4), pectoral muscle (after mastectomy)(26), placenta(1), pooled germ cell tumors(6), prostate(9), senescent fibroblast(6), small cell carcinoma(1), tumor, 5 pooled (see description)(2)
10584	2 pooled tumors (clear cell type)(2), b-cell, chronic lymphocytic leukemia(4), anaplastic oligodendroglioma(6), carcinoid(2), colon(3), colon tumor rer+(4), colon tumor, rer+(2), ovarian tumor(4)
10650	pooled germ cell tumors(2)
10739	2 pooled tumors (clear cell type)(5), ewing's sarcoma(1), liver(1), alveolar rhabdomyosarcoma(1), anaplastic oligodendroglioma(7), breast(3), breast tumor(1), carcinoid(9), colon(5), colon tumor rer+(3), frontal lobe(8), germinal center b cell(3), glioblastoma (pooled)(11), invasive prostate tumor(1), kidney(1), metastatic prostate bone lesion(1), moderately-differentiated adenocarcinoma(1), normal prostate(2), normal prostatic

	epithelial cells(1), parathyroid tumor(4), pectoral muscle (after mastectomy)(6), placenta(1), pooled germ cell tumors(1), tumor(1), tumor, 5 pooled (see description)(1)
10743	2 pooled tumors (clear cell type)(5), ewing's sarcoma(1), liver(1), schwannoma tumor(1), alveolar rhabdomyosarcoma(1), anaplastic oligodendroglioma(7), breast(3), breast tumor(1), carcinoid(10), colon(5), colon tumor rer+(3), germinal center b cell(3), glioblastoma (pooled)(11), invasive prostate tumor(1), kidney(1), liver(1), metastatic prostate bone lesion(1), moderately-differentiated adenocarcinoma(1), muscle(1), normal prostate(2), normal prostatic epithelial cells(1), ovarian tumor(4), parathyroid tumor(4), pectoral muscle (after mastectomy)(6), placenta(1), pooled germ cell tumors(2), thyroid(1), tumor(1), tumor, 5 pooled (see description)(1)
10744	heart(2), four pooled pituitary adenomas(1), frontal lobe(3), glioblastoma (pooled)(1), liver(1), moderately-differentiated adenocarcinoma(1), peripheral blood(1), retina(1)
10880	2 pooled tumors (clear cell type)(3), ewing's sarcoma(1), adenocarcinoma(3), brain(1), carcinoid(8), germinal center b cell(1), glioblastoma (pooled)(5), kidney tumor(1), normal prostate(3), oligodendroglioma(1), parathyroid tumor(11), senescent fibroblast(4), three pooled meningiomas(1), total brain(3)
10942	2 pooled tumors (clear cell type)(8), adenocarcinoma(1), anaplastic oligodendroglioma(1), breast carcinoma in situ(3), cerebellum(1), oligodendroglioma(1), papillary serous carcinoma(2), parathyroid tumor(26), pooled germ cell tumors(1)
11019	epidermis(1), ewing's sarcoma(2), heart(1), schwannoma tumor(1), adenocarcinoma(5), alveolar rhabdomyosarcoma(3), amygdala(1), brain(4), carcinoid(1), colon mucosa(1), endometrioid ovarian metastasis(1), epithelium (cell line)(2), frontal lobe(1), germ cell tumor(3), heart(1), invasive prostate tumor(2), kidney(2), kidney tumor(2), liposarcoma(1), liver(24), mantle cell lymphoma(3), medulloblastoma(1), metastatic prostate bone lesion(6), moderately-differentiated adenocarcinoma(1), muscle(1), normal prostatic epithelial cells(5), ovary(1), papillary serous ovarian metastasis(1), parathyroid tumor(1), placenta(2), pooled germ cell tumors(2), prostate(1), thyroid(1), tumor(1), tumor, 5 pooled (see description)(1), uterus(4)
11278	2 pooled tumors (clear cell type)(4), bone(4), heart(2), anaplastic oligodendroglioma(4), carcinoid(3), colon tumor rer+(1), epithelium (cell line)(1), frontal lobe(12), kidney(2), liposarcoma(1), liver(6), lung carcinoma(1), muscle(6), normal prostate(2), ovary(2), parathyroid tumor(6), pectoral muscle (after mastectomy)(3), schizophrenic brain s-11 frontal lobe(1), senescent fibroblast(6), small cell carcinoma(4), synovial membrane(1)
11342	2 pooled tumors (clear cell type)(3), blood(1), germinal center b cell(4), normal epithelium(1), pooled germ cell tumors(1), tumor, 5 pooled (see description)(1)
11735	adenocarcinoma(1), breast(2), colon(1), frontal lobe(2), placenta(1), pooled germ cell tumors(8)
12039	testis(1)
12043	spleen(1)
12048	2 pooled tumors (clear cell type)(1), anaplastic oligodendroglioma(1), brain(2), colon tumor rer+(2), spleen(2)
12098	astrocytoma(1), ovarian tumor(1)
12202	2 pooled tumors (clear cell type)(2), adenocarcinoma(1), cochlea(1), germinal center b cell(2), pituitary(1)
12243	alveolar rhabdomyosarcoma(2), anaplastic oligodendroglioma(1), carcinoid(9), colon(1), colon tumor rer+(1), germinal center b cell(2), glioblastoma (pooled)(2), metastatic prostate bone lesion(1), normal prostate(2), ovary(2), pooled germ cell tumors(2), thyroid(2), tumor(1)
12263	anaplastic oligodendroglioma(1), carcinoid(1), germinal center b cell(2), invasive adenocarcinoma(1), liver and spleen(2), papillary serous carcinoma(1)
12490	2 pooled tumors (clear cell type)(3), b-cell, chronic lymphocytic leukemia(1), ewing's sarcoma(1), adenocarcinoma(2), anaplastic oligodendroglioma(1), germinal center b cell(2), kidney tumor(1), larynx(1), mantle cell lymphoma(1), medulloblastoma(1), melanocyte(2), moderately-differentiated adenocarcinoma(1), ovarian tumor(12), parathyroid tumor(2), pooled germ cell tumors(4), prostate(2), tumor, 5 pooled (see description)(2)
12604	2 pooled tumors (clear cell type)(12), b-cell, chronic lymphocytic leukemia(4), bone(1), ewing's sarcoma(1), adenocarcinoma(1), anaplastic oligodendroglioma(10), carcinoid(3), colon(3), colon tumor rer+(1), germinal center b cell(1), glioblastoma (pooled)(3), invasive prostate tumor(1), kidney(1), larynx(1), muscle(1), normal leg muscle(1), normal prostate(3), ovarian tumor(1), parathyroid tumor(12), pectoral muscle (after mastectomy)(28), pooled germ cell

	tumors(3), prostate(3), renal cell tumor(1), senescent fibroblast(12), skeletal muscle(1)
12657	adenocarcinoma(1), carcinoid(2), colon tumor, rer+(2)
12788	2 pooled tumors (clear cell type)(3), adenocarcinoma(1), adrenal adenoma(1), alveolar rhabdomyosarcoma(1), carcinoid(1), colon(1), colon tumor rer+(1), germinal center b cell(1), glioblastoma (pooled)(2), ovary(1), parathyroid tumor(1), senescent fibroblast(2), tumor, 5 pooled (see description)(1)
12901	anaplastic oligodendroglioma(2), carcinoid(1), germinal center b cell(6), invasive tumor (cell line)(1), parathyroid tumor(1)
12907	adipose tissue, white(1), carcinoid(1), parathyroid tumor(2), pooled frontal lobe(1), tumor, 5 pooled (see description)(1)
13013	invasive tumor (cell line)(1), parathyroid tumor(3)
13202	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(3), breast(1), breast tumor(1), carcinoid(2), colon tumor rer+(1), colon tumor, rer+(1), germinal center b cell(1), kidney tumor(1), lung carcinoma(1), lung tumor(1), meningioma(2), moderately differentiated adenocarcinoma(1), normal prostate(3), ovarian tumor(1), squamous cell carcinoma(1), tumor, 5 pooled (see description)(1)
13229	cochlea(1), frontal cortex(1), pooled germ cell tumors(1)
13256	2 pooled tumors (clear cell type)(2), adenocarcinoma(2), anaplastic oligodendroglioma(1), parathyroid tumor(1), pooled germ cell tumors(3)
13285	2 pooled tumors (clear cell type)(6), carcinoid(1), colon tumor rer+(1), kidney(1), liver(1)
26710	2 pooled tumors (clear cell type)(1), pooled germ cell tumors(1)
27273	anaplastic oligodendroglioma(1), cerebellum(1), germinal center b cell(2), moderately-differentiated adenocarcinoma(1), pectoral muscle (after mastectomy)(2)
27301	astrocytoma(2), ovarian tumor(1)
27336	2 pooled tumors (clear cell type)(2), b-cell, chronic lymphocytic leukemia(6), adenocarcinoma(1), anaplastic oligodendroglioma(9), carcinoid(6), colon(1), colon tumor rer+(1), colon tumor, rer+(6), frontal lobe(1), germinal center b cell(2), glioblastoma (pooled)(1), pooled germ cell tumors(4), senescent fibroblast(2)
27361	colon(1)
27374	b-cell, chronic lymphocytic leukemia(1), pooled germ cell tumors(5)
27627	2 pooled tumors (clear cell type)(1), anaplastic oligodendroglioma(1), frontal lobe(1), germinal center b cell(1), glioblastoma (pooled)(2), ovarian tumor(1)
27697	skeletal muscle(2), thyroid(1)
27877	carcinoid(1), colon(1), germinal center b cell(1)
29469	frontal lobe(4), germinal center b cell(1), ovarian tumor(2), pectoral muscle (after mastectomy)(3), senescent fibroblast(1), small cell carcinoma(1), thyroid(1)
29557	ewing's sarcoma(1), larynx(1), medulloblastoma(1), moderately-differentiated adenocarcinoma(1), ovarian tumor(7), parathyroid tumor(2), pooled germ cell tumors(1), prostate(1)
29673	heart(1), colon tumor(1), kidney tumor(1), lung tumor(3), lymphoid(2), moderately differentiated adenocarcinoma(1), muscle(1), ovarian tumor(3), squamous cell carcinoma(1)
29814	brain(2), colon(1), pancreatic islet(1)
30218	small cell carcinoma(2)
30446	adenocarcinoma(2), adrenal adenoma(3), colon tumor(1), glioblastoma (pooled)(2), kidney tumor(1), ovarian tumor(1), parathyroid tumor(1), prostate(1), small cell carcinoma(1)
30583	carcinoid(1), ovarian tumor(1), pooled germ cell tumors(2)
30719	2 pooled tumors (clear cell type)(3), b-cell, chronic lymphocytic leukemia(1), breast(2), colon(1), colon tumor, rer+(1), germinal center b cell(1), glioblastoma (pooled)(3), liver(2), normal prostate(2), pooled germ cell tumors(10)
31356	pooled germ cell tumors(1)
31422	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(2), bone(1), anaplastic oligodendroglioma(11), bone marrow stroma(1), brain(2), breast(1), carcinoid(20), colon(6), colon tumor, rer+(1), germinal center b cell(2), glioblastoma (pooled)(3), kidney tumor(1), larynx(1), medulloblastoma(1), normal prostate(1), ovarian tumor(1), parathyroid tumor(1), pectoral muscle (after mastectomy)(2), pooled germ cell tumors(2), senescent fibroblast(5), tumor(1), tumor, 5 pooled (see description)(3)
31554	2 pooled tumors (clear cell type)(1), carcinoid(1), germinal center b cell(4)
31627	2 pooled tumors (clear cell type)(1), alveolar rhabdomyosarcoma(1), anaplastic oligodendroglioma(2), frontal lobe(7), normal prostate(1), oligodendroglioma(1), parathyroid

	tumor(1), pectoral muscle (after mastectomy)(3), pooled germ cell tumors(1)
31744	adenocarcinoma(1)
31790	ovarian tumor(1), synovial membrane(1)
32102	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(2), heart(3), lymphoma(2), adenocarcinoma(6), adipose tissue, white(1), anaplastic oligodendroglioma(5), brain(2), carcinoid(9), cerebellum(5), colon(7), epithelium (cell line)(1), frontal lobe(4), germ cell tumor(1), glioblastoma (pooled)(3), juvenile granulosa tumor(1), kidney tumor(2), medulloblastoma(1), melanoma (mewo cell line)(1), metastatic melanoma to bowel(1), moderately-differentiated adenocarcinoma(1), normal epithelium(1), normal prostate(3), omentum(1), ovarian tumor(2), pancreatic islet(1), parathyroid tumor(4), pectoral muscle (after mastectomy)(2), placenta(1), pooled frontal lobe(1), pooled germ cell tumors(1), retinal fovea(1), schizophrenic brain s-11 frontal lobe(1), senescent fibroblast(5), small cell carcinoma(4), synovial membrane(1), total brain(2), tumor, 5 pooled (see description)(1), two pooled squamous cell carcinomas(2)
32473	2 pooled tumors (clear cell type)(3), heart(1), adipose tissue, white(1), alveolar rhabdomyosarcoma(3), anaplastic oligodendroglioma(3), breast(2), colon(2), colon tumor rer+(1), invasive prostate tumor(1), kidney(1), metastatic prostate bone lesion(1), normal prostate(1), pectoral muscle (after mastectomy)(22), pooled germ cell tumors(2), senescent fibroblast(1)
32475	2 pooled tumors (clear cell type)(3), heart(1), adipose tissue, white(1), alveolar rhabdomyosarcoma(3), anaplastic oligodendroglioma(3), breast(2), colon(2), colon tumor rer+(1), kidney(1), metastatic prostate bone lesion(1), normal prostate(1), pectoral muscle (after mastectomy)(22), pooled germ cell tumors(2), senescent fibroblast(1)
33712	2 pooled tumors (clear cell type)(1), adenocarcinoma(2), anaplastic oligodendroglioma(4), brain frontal cortex(2), breast(1), germinal center b cell(4), juvenile granulosa tumor(1), normal prostatic epithelial cells(2), ovarian tumor(1), three pooled meningiomas(2)
35005	bone(1), pooled germ cell tumors(1)
35185	spinal cord(1)
35326	brain(1), breast(2), carcinoid(1), colon(1), germinal center b cell(2), human pancreatic islets(1), liver(1), melanocyte(1), ovarian tumor(1), placenta(1), total brain(1)
37337	anaplastic oligodendroglioma(2), colon(2), glioblastoma (pooled)(3), juvenile granulosa tumor(1), pooled germ cell tumors(3), tumor, 5 pooled (see description)(1), uterus(2)
38220	2 pooled tumors (clear cell type)(1), b-cell, chronic lymphocytic leukemia(1), adenocarcinoma(2), breast(1), carcinoid(1), frontal lobe(1), germinal center b cell(2), normal prostate(1), normal prostatic epithelial cells(1), pooled germ cell tumors(3), total brain(2)
38311	bone(1), frontal lobe(1), germinal center b cell(1), moderately differentiated adenocarcinoma(1), placenta(1)
38631	brain(1)
38749	2 pooled tumors (clear cell type)(4), breast(1), cochlea(1), colon tumor rer+(1), liver(1), moderately differentiated adenocarcinoma(1), normal prostate(1), pancreas (with no medical abnormalities)(1), parathyroid tumor(1), prostate(1)
40975	brain(2), tumor(1)
40991	brain(2), tumor(1)
44053	2 pooled tumors (clear cell type)(1), bone(1), total brain(1)
45179	colon(1), metastatic prostate bone lesion(1)
45274	2 pooled tumors (clear cell type)(6), cd34+, cd38- from normal bone marrow donor(1), ewing's sarcoma(1), heart(1), lung(1), adenocarcinoma(8), adrenal adenoma(1), alveolar rhabdomyosarcoma(3), anaplastic oligodendroglioma(5), borderline ovarian carcinoma(17), brain(3), breast(4), breast carcinoma in situ(20), bronchioalveolar carcinoma(4), carcinoid(4), cerebellum(1), colon(4), colon tumor rer+(3), early stage papillary serous carcinoma(3), glioblastoma (pooled)(2), invasive adenocarcinoma(8), invasive carcinoma(2), lobular carcinoma in situ(2), low-grade prostatic neoplasia(1), lung carcinoma(1), lung tumor(1), normal prostate(2), normal prostatic epithelial cells(1), oligodendroglioma(3), ovarian tumor(2), papillary serous carcinoma(18), papillary serous ovarian metastasis(1), parathyroid tumor(5), pectoral muscle (after mastectomy)(7), pooled germ cell tumors(2), senescent fibroblast(3), stem cell 34+/38+(2), thymus(1)
46679	2 pooled tumors (clear cell type)(4), anaplastic oligodendroglioma(1), pectoral muscle (after mastectomy)(1), pooled germ cell tumors(1)
48024	2 pooled tumors (clear cell type)(3), heart(1), adipose tissue, white(1), alveolar rhabdomyosarcoma(3), anaplastic oligodendroglioma(3), breast(2), colon(2), colon tumor

	rer+(1), invasive prostate tumor(1), kidney(1), metastatic prostate bone lesion(1), normal prostate(1), pectoral muscle (after mastectomy)(22), pooled germ cell tumors(2), senescent fibroblast(1)
48548	liver(4)
48823	2 pooled tumors (clear cell type)(1), bone(1), alveolar rhabdomyosarcoma(2), breast(1), epithelium (cell line)(1), glioblastoma (pooled)(1), lung(1), lung tumor(1), parathyroid tumor(1), pectoral muscle (after mastectomy)(1), renal cell tumor(1)
48901	germinal center b cell(1), pooled germ cell tumors(1)
49018	anaplastic oligodendroglioma(1), brain(1), breast(1), frontal lobe (see description)(1), testis(5)
49034	adipose tissue, white(1), pancreas (with no medical abnormalities)(1)
49133	retina(1)
49140	2 pooled tumors (clear cell type)(1), adenocarcinoma(1), carcinoid(6), colon tumor rer+(1), glioblastoma (pooled)(2), moderately differentiated adenocarcinoma(1), normal prostate(1), pancreatic islet(2), parathyroid tumor(3)
49387	colon(1), pooled germ cell tumors(1)
49416	anaplastic oligodendroglioma(1), carcinoid(3), cerebellum(2), frontal lobe(2), germinal center b cell(2), parathyroid tumor(1), retina(1), tumor(1)
49493	neuroepithelial cells(1), ovarian tumor(4), thyroid(1)
49640	2 pooled tumors (clear cell type)(1), anaplastic oligodendroglioma(1), brain(2), colon tumor rer+(1), spleen(2)
49863	b-cell, chronic lymphocytic leukemia(1), carcinoid(1), germinal center b cell(2), pooled germ cell tumors(3)
49871	b-cell, chronic lymphocytic leukemia(1), carcinoid(1), germinal center b cell(2), pooled germ cell tumors(3)
50185	total brain(1)
50763	cerebral cortex(3), brain(2), colon(1), colon tumor rer+(1), ovarian tumor(1), schizophrenic brain s-11 frontal lobe(1)
50982	2 pooled tumors (clear cell type)(2), b-cell, chronic lymphocytic leukemia(3), carcinoid(1), cochlea(1), germinal center b cell(2)
51130	adenocarcinoma(1)
51212	placenta.(2)
51346	bone(1), small cell carcinoma(2)
51380	placenta(4)
51400	liver(1)
51954	germinal center b cell(1), total brain(1)
52076	normal prostatic epithelial cells(1)

TABLE V

SEQ ID NO. in priority application	Low frequency expression	High frequency expression
37	salivary gland	
51	fetal brain, fetal kidney	brain, salivary gland, liver
179	liver	
180	liver	
183	prostate	
326	salivary gland	
362	testis	
440	brain	
452	placenta	
483	placenta	
500	salivary gland	
528	testis	

573	liver	
587	salivary gland	
588	liver	
593	fetal kidney, placenta	
599	liver	
628	testis	
653	testis	
670	testis	
693	placenta	
703	liver	
746	brain	
770	placenta	
775	testis	
812	brain	
940	liver	
988	liver	
996	testis	
1036	salivary gland, brain	
1064	brain, liver, prostate	fetal brain
1151	brain	
1190	salivary gland, prostate	
1458	stomach/intestine	
1904	prostate	
2028	brain	
2368	testis	
2553	prostate	
2556	prostate	
2690	testis	
2755	prostate	
2800	liver, testis	
2843	brain	
2932	fetal liver	
3078	prostate	
3280	liver	
3326	prostate	
3387	prostate	
3439	salivary gland	
3501	brain	
3633	brain	
3678	testis	
3714	liver	
3801	testis	
3804	salivary gland, fetal kidney	
3892	brain	
3985	stomach/intestine	
4005	liver	
4063	fetal brain	
4088	testis	
4111	brain	
4126	liver	
4172	liver	
4261	prostate	
4436	liver	
4609	fetal liver	
4647	fetal liver	
4660	fetal liver	

4664	fetal brain	
4678	fetal liver	
4687	fetal liver	
4690	fetal liver	
4694	salivary gland	
4696	salivary gland	
4733	fetal liver	
4807	fetal brain	
4809	salivary gland	
4855	prostate	
4908	brain	
4943	liver	
4947	testis	
4976	stomach/intestine	
5000	fetal kidney	
5005	fetal kidney	
5011	prostate	
5040	fetal liver	
5089	salivary gland	
5117	prostate	
5141	liver	
5162	fetal brain	
5167	salivary gland	
5192	liver	
5214	fetal brain	
5250	fetal kidney, fetal brain	
5262	brain	fetal kidney, fetal brain
5270	fetal brain	
5278	fetal liver	
5453	fetal kidney	
5494	fetal liver	
5499	brain	
5533	fetal brain	
5563	fetal liver	
5609	placenta	
5748	placenta	
5816	fetal brain	
5824	prostate	
5861	fetal liver, prostate, fetal kidney, salivary gland	
5947	fetal liver	
5966	placenta	
5970	fetal kidney	
5974	fetal liver	
5983	brain	
5985	fetal kidney	
6080	salivary gland	
6081	salivary gland	
6108	liver	
6238	placenta, brain	
6252	placenta	
6334	liver	
6345	brain	
6358	fetal brain	
6636	fetal brain	
6688	fetal brain	

6727	testis	
6865	salivary gland	
6892	stomach/intestine, fetal brain	
7041	liver	
7533	placenta	
7535	brain	
7697	fetal liver	
7712	testis	
8078	testis	
8097	salivary gland	
8166	placenta, fetal liver	
8262	prostate	
8534	testis	
8666	placenta	
8744	prostate	
8968	brain	
8994	testis	
9297	fetal liver	
9332	testis	
9406	prostate	
9407	prostate	
9668	testis	
9755	testis	
9868	liver	
10044	fetal liver, placenta	
10526	prostate	
10650	testis	
10743	stomach/intestine	
10744	salivary gland	
10761	prostate	
10880	fetal liver	
10942	brain	
11019	liver	
11278	salivary gland	
11562	testis	
11688	testis	
11735	fetal brain	
11813	brain	
12043	brain	
12202	testis	
12220	fetal kidney	
12243	brain, testis, liver	fetal brain, fetal liver
12263	prostate	
12276	brain	
12604	fetal kidney	
12657	fetal kidney, stomach/intestine	
12788	testis	
12901	placenta	
12907	salivary gland	
13013	fetal liver	
13229	brain	
13256	testis	
13267	salivary gland	
13285	fetal brain	fetal kidney, stomach/intestine
26638	fetal liver	
26710	brain	

26726	fetal kidney	
26786	brain	
27084	testis	
27273	fetal kidney	
27361	liver	
27374	fetal liver	
27877	salivary gland	
28413	prostate	
28517	testis	
28518	placenta	
29120	salivary gland	
29673	fetal liver	
30218	liver	
30446	prostate	
30477	placenta	
30583	testis	
30719	fetal brain, salivary gland	
31356	prostate	
31790	brain	
33130	testis	
38749	prostate	
38890	placenta	
40163	testis	
40975	prostate	
40991	prostate	
42896	prostate	
44053	stomach/intestine, fetal kidney	
45091	stomach/intestine	
45179	prostate	
45274	fetal kidney	
46679	fetal liver	
48024	fetal kidney	
48603	liver, testis	
48671	testis	
48823	fetal kidney	
48901	salivary gland	
49018	stomach/intestine	
49034	salivary gland	
49133	liver	
49261	testis	
49387	liver	
49416	fetal brain	
49426	stomach/intestine	
49493	salivary gland	
49640	brain	
49863	brain	
49871	brain	
50015	testis	
50049	testis	
50112	testis	
50185	salivary gland	
50241	liver	
50763	brain	
51130	testis	
51212	stomach/intestine	
51400	liver	

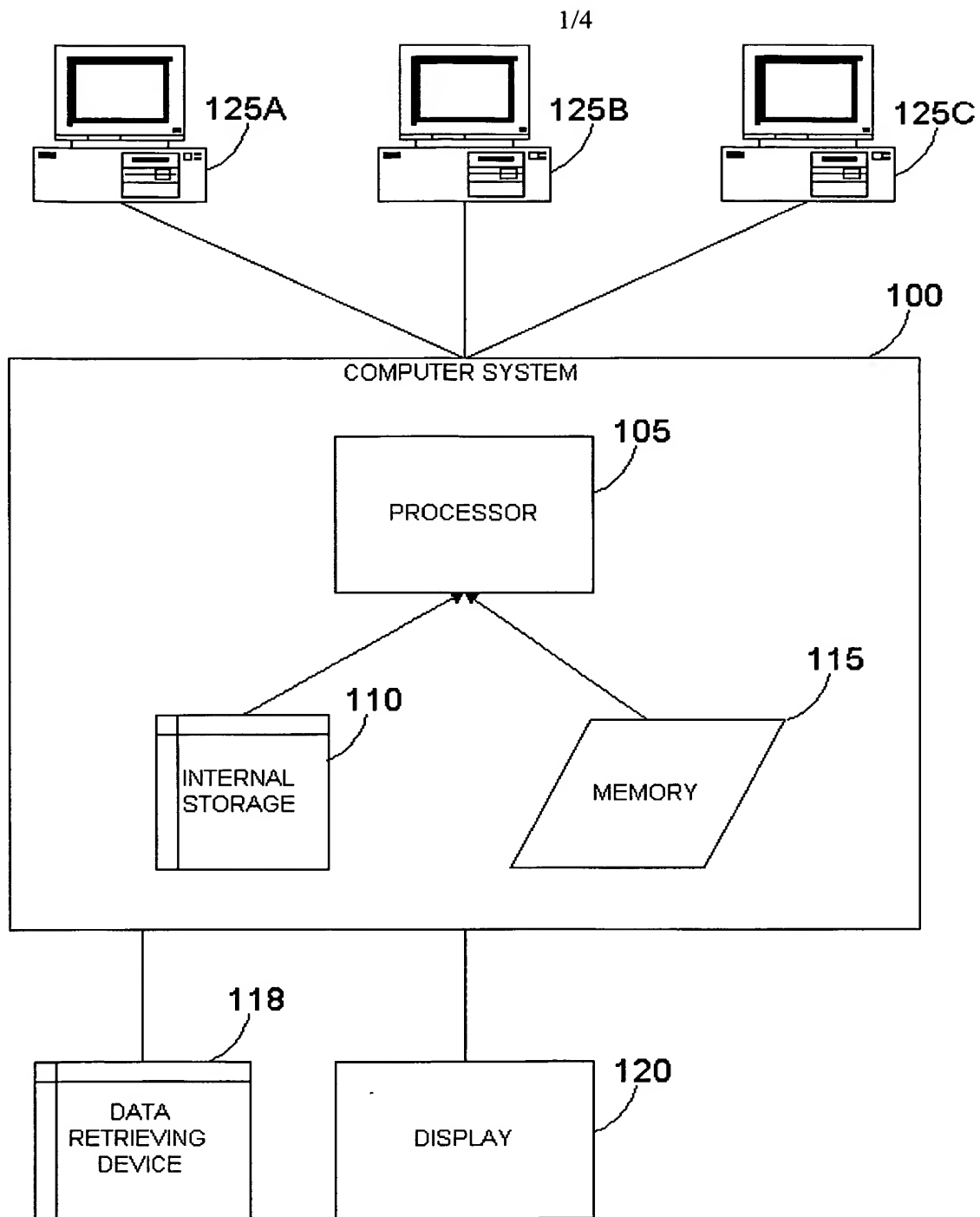
51954	fetal brain	
-------	-------------	--

WHAT IS CLAIMED IS:

1. An isolated polynucleotide, said polynucleotide comprising a nucleic acid sequence encoding:
 - i) A polypeptide comprising the amino acid sequence shown as SEQ
5 ID NO:305;
 - ii) a polypeptide comprising any one of the amino acid sequences
shown as SEQ ID NOs:170-304, 306-338, 456-560, 785-918; or
 - iii) a biologically active fragment of any of said polypeptides.
- 10 2. The polynucleotide of claim 1, wherein said polypeptide comprises a signal peptide.
3. The polynucleotide of claim 1, wherein said polypeptide is a mature protein.
4. The polynucleotide of claim 1, wherein said polynucleotide comprises any one of the nucleic
15 acid sequences shown as SEQ ID NOs:1-169, 339-455, or 561-784.
5. The polynucleotide of claim 1, wherein said polynucleotide is operably linked to a promoter.
6. An expression vector comprising the polynucleotide of claim 5.
20
7. A host cell recombinant for the polynucleotide of claim 1.
8. A non-human transgenic animal comprising the host cell of claim 7.
- 25 9. A pharmaceutical composition comprising the polynucleotide of claim 1, and a
pharmaceutically acceptable carrier.
10. A method of making a GENSET polypeptide, said method comprising
 - a) providing a population of host cells comprising the polynucleotide of
30 claim 5; and
 - b) culturing said population of host cells under conditions conducive to the
production of said polypeptide within said host cells.
11. The method of claim 10, further comprising purifying said polypeptide from said population of
35 host cells.

12. An isolated polynucleotide, said polynucleotide comprising any one of the nucleic acid sequences shown as SEQ ID NOs:1-169, 339-455, or 561-784.
13. A biologically active polypeptide encoded by the polynucleotide of claim 12.
14. An isolated polypeptide or biologically active fragment thereof, said polypeptide comprising any one of the amino acid sequences shown as SEQ ID NOs:170-338, 456-560, or 785-918.
15. The polypeptide of claim 14, wherein said polypeptide comprises a signal peptide.
16. The polypeptide of claim 14, wherein said polypeptide is a mature protein.
17. An antibody that specifically binds to the polypeptide of claim 14.
18. A pharmaceutical composition comprising the polypeptide of claim 14, and a pharmaceutically acceptable carrier.
19. A method of making a GENSET polypeptide, said method comprising
- a) providing a population of cells comprising a polynucleotide encoding the polypeptide of claim 14, operably linked to a promoter;
 - b) culturing said population of cells under conditions conducive to the production of said polypeptide within said cells; and
 - c) purifying said polypeptide from said population of cells.
20. A method of determining whether a GENSET gene is expressed within a mammal, said method comprising the steps of:
- a) providing a biological sample from said mammal
 - b) contacting said biological sample with either of:
 - i) a polynucleotide that hybridizes under stringent conditions to the polynucleotide of claim 1; or
 - ii) a polypeptide that specifically binds to the polypeptide of claim 14; and
 - c) detecting the presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said polypeptide to a protein within said sample;
- wherein a detection of said hybridization or of said binding indicates that said GENSET gene is expressed within said mammal.

21. The method of claim 21, wherein said polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising the sequence of said primer.
- 5 22. The method of claim 21, wherein said polypeptide is an antibody.
23. A method of determining whether a mammal has an elevated or reduced level of GENSET gene expression, said method comprising the steps of :
- 10 a) providing a biological sample from said mammal; and
- b) comparing the amount of the polypeptide of claim 14, or of an RNA species encoding said polypeptide, within said biological sample with a level detected in or expected from a control sample;
- wherein an increased amount of said polypeptide or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said
- 15 mammal has an elevated level of said GENSET gene expression, and wherein a decreased amount of said polypeptide or said RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level of said GENSET gene expression.
- 20 24. A method of identifying a candidate modulator of a GENSET polypeptide, said method comprising :
- a) contacting the polypeptide of claim 14 with a test compound; and
- b) determining whether said compound specifically binds to said polypeptide;
- 25 wherein a detection that said compound specifically binds to said polypeptide indicates that said compound is a candidate modulator of said GENSET polypeptide.
25. The method of claim 24, further comprising testing the biological activity of said GENSET polypeptide in the presence of said candidate modulator, wherein an alteration in the biological
- 30 activity of said GENSET polypeptide in the presence of said compound in comparison to the activity in the absence of said compound indicates that the compound is a modulator of said GENSET polypeptide.
26. A method for the production of a pharmaceutical composition comprising
- 35 a) identifying a modulator of a GENSET polypeptide using the method of claim 24; and
- b) combining said modulator with a pharmaceutically acceptable carrier.

**Figure 1**

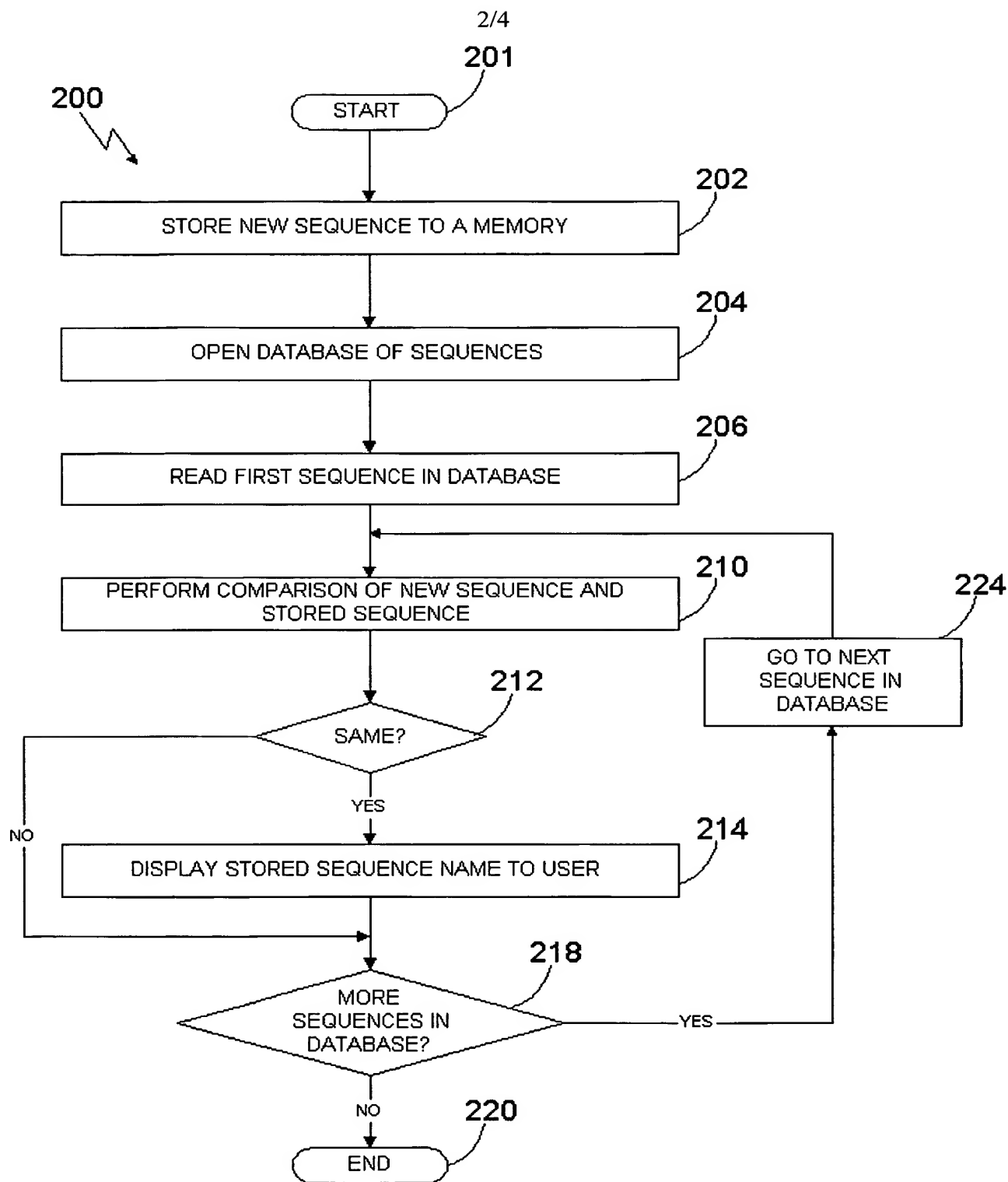
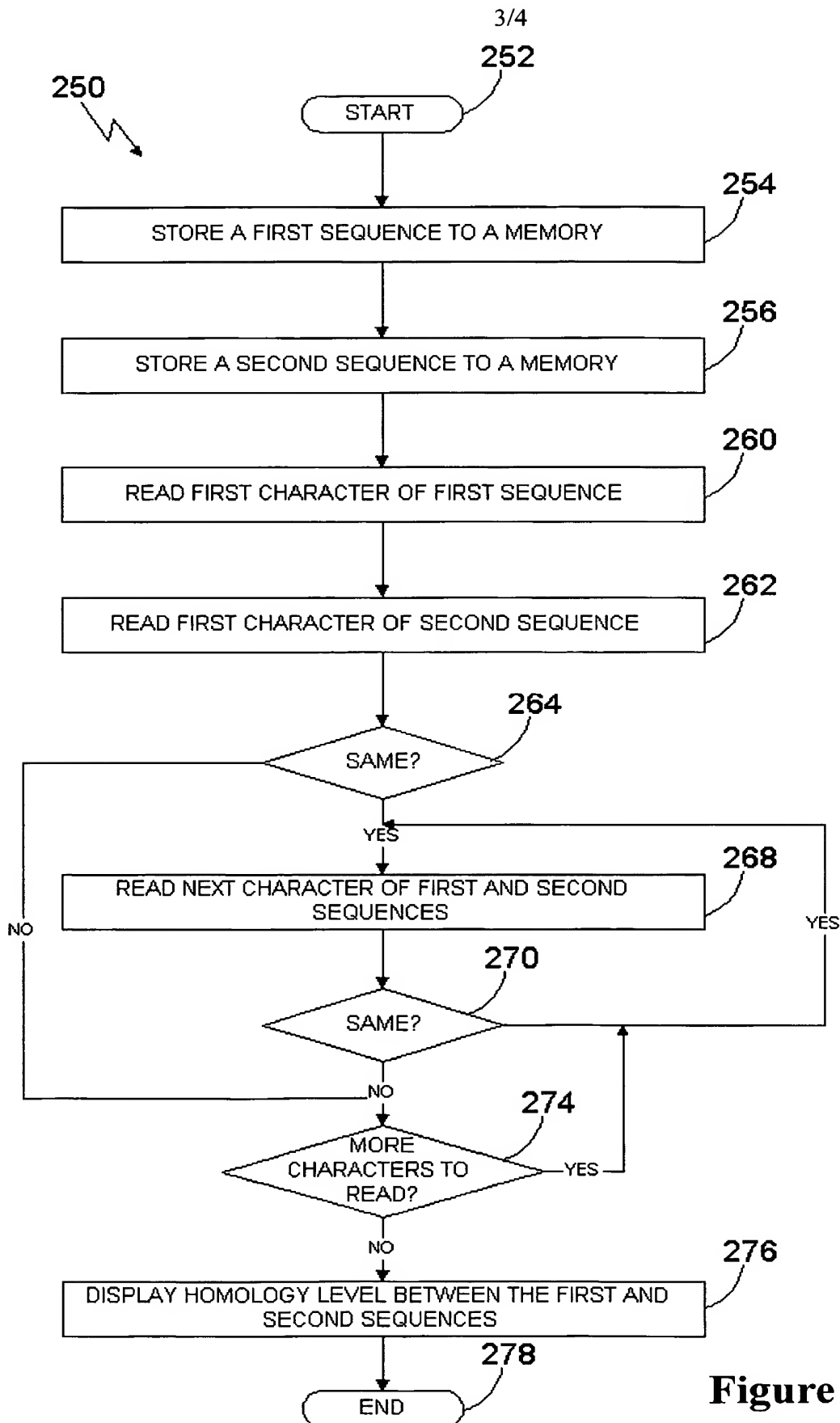
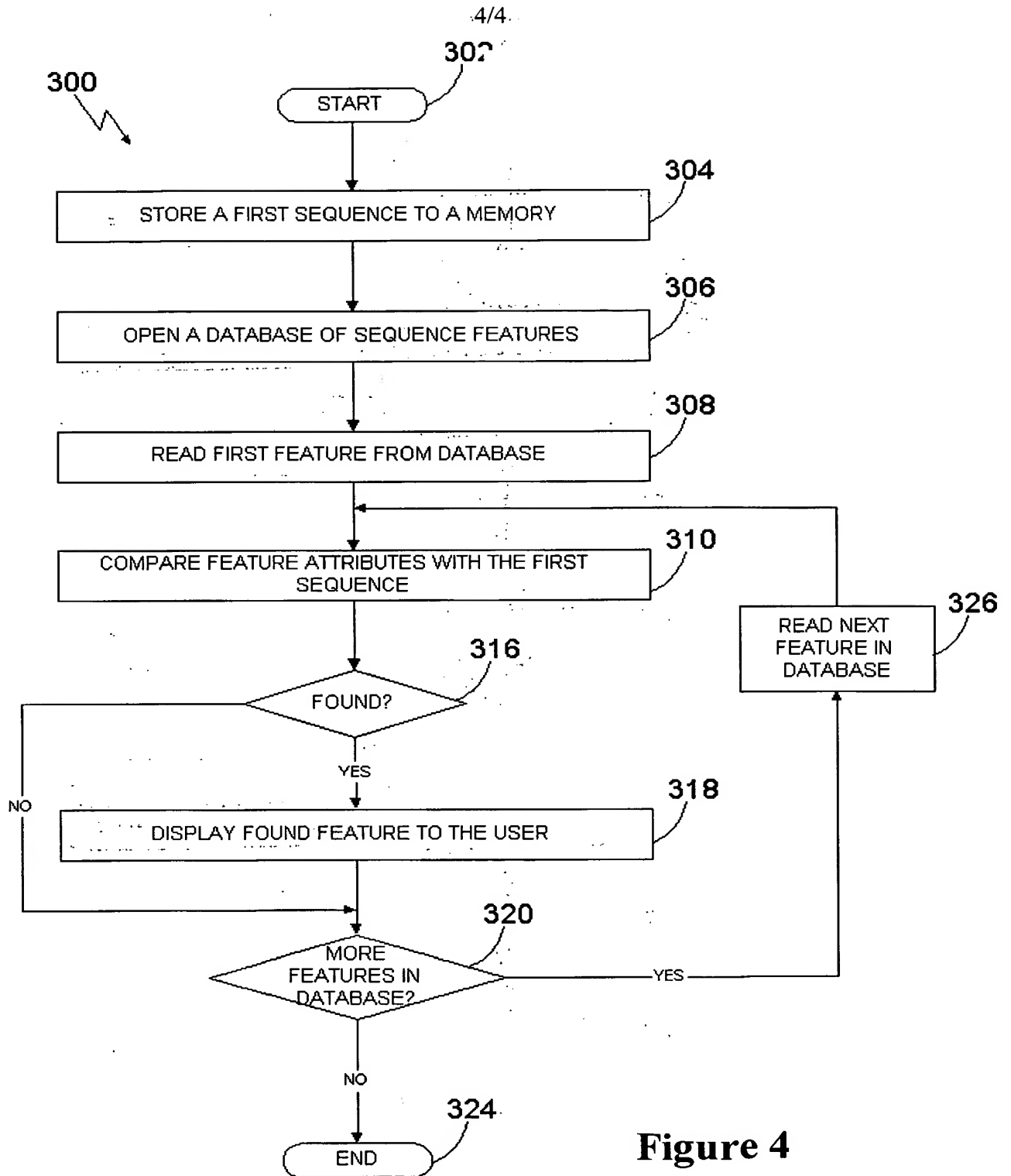


Figure 2

**Figure 3**

**Figure 4**

SEQUENCE LISTING

<110> GENSET SA

<120> Full-length human cDNAs encoding potentially secreted proteins

<130> 81.WO

<150> US 60/197,873

<151> 2000-04-18

<150> US 60/224,009

<151> 2000-08-07

<150> US 60/260,328

<151> 2001-01-08

<150> US 60/224,006

<151> 2000-08-04

<160> 918

<170> Patent.pm

<210> 1

<211> 615

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 88..471

<220>

<221> sig_peptide

<222> 88..189

<223> Von Heijne matrix

score 13.1278588169685

seq VLLLLSLLLTAXA/SS

<220>

<221> polyA_site

<222> 600..615

<220>

<221> polyA_signal

<222> 577..582

<400> 1

acttatctgc agactttagt gcagcaactc accctcactc agaggtcttc tggttctgga 60

aacaactcta gctcagcctt ctccacc atg agc ctc aga ctt gat acc acc cct 114

Met Ser Leu Arg Leu Asp Thr Thr Pro

-30

tcc tgt aac agt gcg aga cca ctt cat gcc ttg cag gtg ctg ctg ctt 162

Ser Cys Asn Ser Ala Arg Pro Leu His Ala Leu Gln Val Leu Leu Leu

-25 -20 -15 -10

ctg tca ttg ctg ctg act gct ytg gct tcc tcc acc aaa gga caa ayt 210

Leu Ser Leu Leu Leu Thr Ala Leu Ala Ser Ser Thr Lys Gly Gln Xaa

2

```

      -5              1              5
aag aga aac ttg gcg aaa ggc aaa gac gaa agt cta gac agt gac ttg      258
Lys Arg Asn Leu Ala Lys Gly Lys Asp Glu Ser Leu Asp Ser Asp Leu
      10              15              20
tat gct gaa ctg cgc tgc atg tgt ata aag aca acc tct gga att cat      306
Tyr Ala Glu Leu Arg Cys Met Cys Ile Lys Thr Thr Ser Gly Ile His
      25              30              35
ccc aaa aac atc caa agt ttg gaa gtg atc ggg aaa gga acc cat tgc      354
Pro Lys Asn Ile Gln Ser Leu Glu Val Ile Gly Lys Gly Thr His Cys
      40              45              50
aac caa gtc gaa gtg ata gcc aca ctg aag gat ggg agg aaa atc tgc      402
Asn Gln Val Glu Val Ile Ala Thr Leu Lys Asp Gly Arg Lys Ile Cys
      60              65              70
ctg gac cca gat gct ccc aga atc aag aaa att gta cag aaa aaa ttg      450
Leu Asp Pro Asp Ala Pro Arg Ile Lys Lys Ile Val Gln Lys Lys Leu
      75              80              85
gca ggt gat gaa tct gct gat taatttggtc tgtttctgcc aaacttcttt      501
Ala Gly Asp Glu Ser Ala Asp
      90
aactcccagg aagggtagaa ttttgaaacc ttgattttct agttctcatt tattcaggat      561
acctattctt actgtattaa aatttgata tgtgtttcaa aaaaaaaaaa aaaa      615

```

<210> 2

<211> 539

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 39..518

<220>

<221> sig_peptide

<222> 39..92

```

<223> Von Heijne matrix
      score 12.2866240999866
      seq LTLAVLFLTGSQA/RH

```

<220>

<221> polyA_site

<222> 524..539

<220>

<221> polyA_signal

<222> 507..512

<400> 2

```

agagactgcg agaaggaggt ccccccacggc ccttcagg atg aaa gct gcg gtg ctg      56
                               Met Lys Ala Ala Val Leu
                               -15
acc ttg gcc gtg ctg ttc ctg acg ggg agc cag gct cgg cat ttc tgg      104
Thr Leu Ala Val Leu Phe Leu Thr Gly Ser Gln Ala Arg His Phe Trp
      -10              -5              1
cag caa gat gaa ccc ccc cag agc ccc tgg gat cga gtg aag gac ctg      152
Gln Gln Asp Glu Pro Pro Gln Ser Pro Trp Asp Arg Val Lys Asp Leu
      5              10              15              20
gcc act gtg tac gtg gat gtg ctg aaa gac agc ggc aga gac tat gtg      200
Ala Thr Val Tyr Val Asp Val Leu Lys Asp Ser Gly Arg Asp Tyr Val
      25              30              35
tcc cag ttt gaa ggc tcc gcc ttg gga aaa cag cta aac cta aag ctg      248
Ser Gln Phe Glu Gly Ser Ala Leu Gly Lys Gln Leu Asn Leu Lys Leu
      40              45              50
ctt gac aac tgg gac agc gtg acc tcc gcc ttc agc aat ctg cgc gaa      296

```

3

Leu	Asp	Asn	Trp	Asp	Ser	Val	Thr	Ser	Ala	Phe	Ser	Asn	Leu	Arg	Glu		
		55					60					65					
cag	ctc	ggc	cct	gtg	acc	cag	gag	tty	tgg	gat	aac	ctg	gaa	aag	gag	344	
Gln	Leu	Gly	Pro	Val	Thr	Gln	Glu	Phe	Trp	Asp	Asn	Leu	Glu	Lys	Glu		
		70					75				80						
aca	gag	ggc	ctg	agg	cag	gag	atg	agc	aag	gat	ctg	gag	gag	gtg	aag	392	
Thr	Glu	Gly	Leu	Arg	Gln	Glu	Met	Ser	Lys	Asp	Leu	Glu	Glu	Val	Lys		
85					90					95					100		
gcc	aag	gtg	cag	ccc	tac	ctg	gac	gac	ttc	cag	aag	aag	tgg	cag	gag	440	
Ala	Lys	Val	Gln	Pro	Tyr	Leu	Asp	Asp	Phe	Gln	Lys	Lys	Trp	Gln	Glu		
				105					110					115			
gag	atg	gag	ctc	tac	cgc	cag	aag	gca	gct	ttc	tta	act	atc	cta	aca	488	
Glu	Met	Glu	Leu	Tyr	Arg	Gln	Lys	Ala	Ala	Phe	Leu	Thr	Ile	Leu	Thr		
				120				125					130				
agc	ctt	gga	cca	aat	gga	aat	aaa	gct	ttt	tgatgaaaaa	aaaaaaaaaa	a				539	
Ser	Leu	Gly	Pro	Asn	Gly	Asn	Lys	Ala	Phe								
		135					140										

<210> 3

<211> 796

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 134..373

<220>

<221> sig_peptide

<222> 134..211

<223> Von Heijne matrix

score 12.0462155590967

seq LLLALLLPTQIYS/SE

<400> 3

atttgtggct	ttcctggtat	ataaggtctc	gccggctcgc	cgcgctcccc	accttgcttg		60
cgccgcgccg	gagccagcgg	ttctccaagc	acccagcatc	ctgctagacg	cgccgcgcac		120
cgacggaggg	gac atg ggc aga gca atg gtg gcc agg ctc ggg ctg ggg						169
	Met Gly Arg Ala Met Val Ala Arg Leu Gly Leu Gly						
	-25		-20		-15		
ctg ctg ctg ctg gca ctg ctc cta ccc acg cag att tat tcc agt gaa							217
Leu Leu Leu Leu Ala Leu Leu Leu Pro Thr Gln Ile Tyr Ser Ser Glu							
	-10		-5		1		
act aca act gga act tca agt aac tcc tcc cag agt act tcc aac tct							265
Thr Thr Thr Gly Thr Ser Ser Asn Ser Ser Gln Ser Thr Ser Asn Ser							
	5		10		15		
ggg ttg gcc cca aat cca act aat gcc acc acc aag gtg gct ggt ggt							313
Gly Leu Ala Pro Asn Pro Thr Asn Ala Thr Thr Lys Val Ala Gly Gly							
	20		25		30		
gcc ctg cag tca aca gcc agt ctc ttc gtg gtc tca ctc tct ctt ctg							361
Ala Leu Gln Ser Thr Ala Ser Leu Phe Val Val Ser Leu Ser Leu Leu							
	35		40		45		50
cay cyc tac tct taagagactc aggccaagaa acgtyttcta aatttcccca							413
His Xaa Tyr Ser							
tcttctaacc ccaatccaaa tggcgtctgg aagtccaatg tggcaaggaa aaacaggtct							473
tcatcgaatc tactaattcc acacctttta ttgacacaga aaatgttgag aatcccaa							533
ttgattgatt tgaagaacat gtgagagggt tgactagatg atggatgcca atattaa							593
tgctggagtt tcatgtacaa gatgaaggag aggcaacatc caaaatagtt aagacatgat							653
ttccttgaat gtggcttgag aaatatggac acttaatact accttgaaaa taagaataga							713
aataaaggat gggattgtgg aatggagatt cagttttcat ttggttcatt aattctataa							773
ggccataaaa caggtaatat aaa							796

<210> 4

<211> 528
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 75..377

<220>
 <221> sig_peptide
 <222> 75..155
 <223> Von Heijne matrix
 score 11.8486806626293
 LFLGLLLLPLVVA/FA

<220>
 <221> polyA_site
 <222> 513..528

<220>
 <221> polyA_signal
 <222> 498..503

<400> 4
 attggccaca gagaccacgc ccgagtttcc catcgcaactg agcaactgaga tcctgctgga 60
 agctctgccc cagc atg agc tcc gca gcc ggg ttc tgc gcc tca cgc ccc 110
 Met Ser Ser Ala Ala Gly Phe Cys Ala Ser Arg Pro
 -25 -20
 ggg ctg ctg ttc ctg ggg ttg ctg ctc ctg cca ctt gtg gtc gcc ttc 158
 Gly Leu Leu Phe Leu Gly Leu Leu Leu Leu Pro Leu Val Val Ala Phe
 -15 -10 -5 1
 gcc agc gct gaa gct gaa gaa gat ggg gac ctg cag tgc ctg tgt gtg 206
 Ala Ser Ala Glu Ala Glu Glu Asp Gly Asp Leu Gln Cys Leu Cys Val
 5 10 15
 aag acc acc tcc cag gtc cgt ccc agg cac atc acc agc ctg gag gtg 254
 Lys Thr Thr Ser Gln Val Arg Pro Arg His Ile Thr Ser Leu Glu Val
 20 25 30
 atc aag gcc gga ccc cac tgc ccc act gcc caa ctg ata gcc acg ctg 302
 Ile Lys Ala Gly Pro His Cys Pro Thr Ala Gln Leu Ile Ala Thr Leu
 35 40 45
 aag aat gga agg aaa att tgc ttg gac ctg caa gcc ccg ctg tac aag 350
 Lys Asn Gly Arg Lys Ile Cys Leu Asp Leu Gln Ala Pro Leu Tyr Lys
 50 55 60 65
 aaa ata att aag aaa ctt ttg gag agt tagctactag ctgcctacgt 397
 Lys Ile Ile Lys Lys Leu Leu Glu Ser
 70
 gtgtgcattt gctatatagc ataacttcttt tttccagttt caatctaact gtgaaagaac 457
 ttctgatatt tgtgttatcc ttatgatttt aaataaacia aataaatcaa gttgcaaaaa 517
 aaaaaaaaaa a 528

<210> 5
 <211> 522
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 76..378

<220>
 <221> sig_peptide
 <222> 76..156
 <223> Von Heijne matrix

score 11.8486806626293
LFLGLLLLPLVVA/FA

<220>

<221> polyA_site

<222> 508..522

<220>

<221> polyA_signal

<222> 499..504

<400> 5

```

aattggccac agagacccag cccgagtttc ccatcgcaact gagcactgag atcctgctgg      60
aagctctgcc gcagc atg agc tcc gca gcc ggg ttc tgc gcc tca cgc ccc      111
                Met Ser Ser Ala Ala Gly Phe Cys Ala Ser Arg Pro
                -25                                -20

ggg ctg ctg ttc ctg ggg ttg ctg ctc ctg cca ctt gtg gtc gcc ttc      159
Gly Leu Leu Phe Leu Gly Leu Leu Leu Leu Pro Leu Val Val Ala Phe
-15                                -10                                -5                                1
gcc agc gct gaa gct gaa gaa gat ggg gac ctg cag tgc ctg tgt gtg      207
Ala Ser Ala Glu Ala Glu Glu Asp Gly Asp Leu Gln Cys Leu Cys Val
                5                                10                                15

aag acc acc tcc cag gtc cgt ccc agg cac atc acc agc ctg gag gtg      255
Lys Thr Thr Ser Gln Val Arg Pro Arg His Ile Thr Ser Leu Glu Val
                20                                25                                30

atc aag gcc gga ccc cac tgc ccc act gcc caa ctg ata gcc acg ctg      303
Ile Lys Ala Gly Pro His Cys Pro Thr Ala Gln Leu Ile Ala Thr Leu
                35                                40                                45

aag aat gga agg aaa att tgc ttg gac ctg caa gcc ccg ctg tac aag      351
Lys Asn Gly Arg Lys Ile Cys Leu Asp Leu Gln Ala Pro Leu Tyr Lys
50                                55                                60                                65

aaa ata att aag aaa ctt ttg gag agt tagctactag ctgcctacgt      398
Lys Ile Ile Lys Lys Leu Leu Glu Ser
                70

gtgtgcattt gctatatagc atacttcttt tttccagttt caatctaact gtgaaagaac      458
ttctgatatt tgtgttatcc ttatgatttt aaataaacia aataaatcta aaaaaaaaaa      518
aaaa                                                                522

```

<210> 6

<211> 652

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 75..332

<220>

<221> sig_peptide

<222> 75..128

<223> Von Heijne matrix

score 11.7816699306029

seq VLLLLLLVEQAAA/LG

<220>

<221> polyA_site

<222> 637..652

<220>

<221> polyA_signal

<222> 616..621

<400> 6

```

gactcttttgg tcaggggaact gctcgtctgag cacagctgca cagtgtctgtc agaacggccg      60
atctccagcc caag atg att cca gca gtg gtc ttg ctc tta ctc ctt ttg      110
                Met Ile Pro Ala Val Val Leu Leu Leu Leu Leu Leu
                -15                -10

gtt gaa caa gca gcg gcc ctg gga gag cct cag ctc tgc tat atc ctg      158
Val Glu Gln Ala Ala Ala Leu Gly Glu Pro Gln Leu Cys Tyr Ile Leu
-5                1                5                10

gat gcc atc ctg ttt ctg tat gga att gtc ctc acc ctc ctc tac tgt      206
Asp Ala Ile Leu Phe Leu Tyr Gly Ile Val Leu Thr Leu Leu Tyr Cys
                15                20                25

cga ctg aag atc caa gtg cga aag gca gct ata acc agc tat gag aaa      254
Arg Leu Lys Ile Gln Val Arg Lys Ala Ala Ile Thr Ser Tyr Glu Lys
                30                35                40

tca gat ggt gtt tac acg ggc ctg agc acc agg aac cag gag act tac      302
Ser Asp Gly Val Tyr Thr Gly Leu Ser Thr Arg Asn Gln Glu Thr Tyr
                45                50                55

gag act ctg aag cat gag aaa cca ccm cag tagccttaga atagatgcgg      352
Glu Thr Leu Lys His Glu Lys Pro Pro Gln
        60                65

tcatattctt ctttggcttc tggttctctcc agccctcatg tttggcatca catatgcctg      412
catgccatta acaccagctg gccctacccc tataatgata ctgtgtccta aattaatata      472
caccagtggg tctctctccc tggttaaagac taatgctcag atgctgttta cggatattta      532
tattctagtc tcaactctct gtcccaccct tcttctcttc cccattccca actccagcta      592
aatatggga agggagaacc cccaataaaaa ctgccatgga ctggaaaaaa aaaaaaaaaa      652

```

<210> 7

<211> 743

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 112..387

<220>

<221> sig_peptide

<222> 112..183

<223> Von Heijne matrix

score 11.7298925418815

seq FVLGLGLTPPTLA/QD

<220>

<221> polyA_site

<222> 728..743

<220>

<221> polyA_signal

<222> 705..710

<400> 7

```

acacaactgg aacccatctc caggaacaaa cagctggaac ccatctcccg ttgaaggga      60
actgccagat ttttgtaaga ttcttcctcc tgggagcctg tggttggaaga g atg gtg      117
                Met Val

atg ggc ctg ggc gtt ttg ttg ttg gtc ttc gtg ctg ggt ctg ggt ctg      165
Met Gly Leu Gly Val Leu Leu Leu Val Phe Val Leu Gly Leu Gly Leu
-20                -15                -10

acc cca ccg acc ctg gct cag gat aac tcc agg tac aca cac ttc ctg      213
Thr Pro Pro Thr Leu Ala Gln Asp Asn Ser Arg Tyr Thr His Phe Leu
-5                1                5                10

acc cag cac tat gat gcc aaa cca cag ggc cgg gat gac aga tac tgt      261
Thr Gln His Tyr Asp Ala Lys Pro Gln Gly Arg Asp Asp Arg Tyr Cys
                15                20                25

gaa agc atc atg agg aga cgg ggc ctg acc tca ccc tgc aaa gac atc      309

```

```

Glu Ser Ile Met Arg Arg Arg Gly Leu Thr Ser Pro Cys Lys Asp Ile
      30              35              40
aac aca ttt att cat ggc aac aag cgc asa tca agg cca tct gtg aaa      357
Asn Thr Phe Ile His Gly Asn Lys Arg Xaa Ser Arg Pro Ser Val Lys
      45              50              55
aca aga atg gaa acc ctc aca gag aaa acc taagaataag caagtcttct      407
Thr Arg Met Glu Thr Leu Thr Glu Lys Thr
      60              65
ttccaggtca ccacttgcaa gctacatgga ggttccccct ggcctccatg ccagtaccga      467
gccacagcgg gggttcagaaa cgttggtgtt gcttggtgaaa atggcttacc tgtccacttg      527
gatacagtc aaatccggtcg tccgtaacca gcgggcccct ggtcaagtgc tggctctgct      587
gtccttgcc tccatttccc ctctgcaccc agaacagtgg tggcaacatt cattgccaaag      647
ggcccaaaga aagagctacc tggacctttt gttttctgtt tgacaacatg tttaataaat      707
aaaaatgtct tgatatcagc aaaaaaaaaa aaaaaa      743

```

<210> 8
 <211> 502
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 39..371

<220>
 <221> sig_peptide
 <222> 39..86
 <223> Von Heijne matrix
 score 11.56948973398
 seq ILLSVALLAFSSA/QD

<220>
 <221> polyA_site
 <222> 487..502

<220>
 <221> polyA_signal
 <222> 467..472

```

<400> 8
agcataaaagt tgggagtgac accagagcct tctgcaag atg ctt ctg att ctg ctg      56
                        Met Leu Leu Ile Leu Leu
                        -15
tca gtg gcc ctg ctg gcc ttc agc tca gct cag gac tta gat gaa gat      104
Ser Val Ala Leu Leu Ala Phe Ser Ser Ala Gln Asp Leu Asp Glu Asp
-10              -5              1              5
gtc agc caa gaa gac gtt ccc ttg gta ata tca gat gga gga gac tct      152
Val Ser Gln Glu Asp Val Pro Leu Val Ile Ser Asp Gly Gly Asp Ser
      10              15              20
gag cag ttc ata gat gag gag cgt cag gga cca cct ttg gga gga cag      200
Glu Gln Phe Ile Asp Glu Glu Arg Gln Gly Pro Pro Leu Gly Gly Gln
      25              30              35
caa tct caa ccc tct gct ggt gat ggg aac cag aat gat ggc cct cag      248
Gln Ser Gln Pro Ser Ala Gly Asp Gly Asn Gln Asn Asp Gly Pro Gln
      40              45              50
cag gga cca ccc caa caa gga ggc cag cag caa caa ggt cca cca cct      296
Gln Gly Pro Pro Gln Gln Gly Gly Gln Gln Gln Gln Gly Pro Pro Pro
      55              60              65              70
cct cag gga aag cca caa gga cca cct ccc caa ggg ggc cgc cca caa      344
Pro Gln Gly Lys Pro Gln Gly Pro Pro Pro Gln Gly Gly Arg Pro Gln
      75              80              85
gga cct cca cag ggg cag tct cct cag taatctagga ttcaatgaca      391
Gly Pro Pro Gln Gly Gln Ser Pro Gln

```


90 95
ggaagtgaat aagaagatga cagtgtttca aatgccttga aacataatgt gatcatgctc 451
taacttcaat ataccaataa aataatcagc ttgctaaaaa aaaaaaaaaa a 502

<210> 9
<211> 401
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 36..353

<220>
<221> sig_peptide
<222> 36..83
<223> Von Heijne matrix
score 11.56948973398
seq ILLSVALLAFSSA/QD

<220>
<221> polyA_site
<222> 386..401

<400> 9
ataaagttag gagtgacacc agagccttct gcaag atg ctt ctg att ctg ctg 53
Met Leu Leu Ile Leu Leu
-15

tca gtg gcc ctg ctg gcc ttc agc tca gct cag gac tta gat gaa gat 101
Ser Val Ala Leu Leu Ala Phe Ser Ser Ala Gln Asp Leu Asp Glu Asp
-10 -5 1 5

gtc agc caa gaa gac gtt ccc ttg gta ata tca gat gga gga gac tct 149
Val Ser Gln Glu Asp Val Pro Leu Val Ile Ser Asp Gly Gly Asp Ser
10 15 20

gag cag ttc ata gat gag gag cgt cag gga cca cct ttg gga gga cag 197
Glu Gln Phe Ile Asp Glu Glu Arg Gln Gly Pro Pro Leu Gly Gly Gln
25 30 35

caa tct caa ccc tct gct ggt gat ggg aac cag gat gat ggc cct cag 245
Gln Ser Gln Pro Ser Ala Gly Asp Gly Asn Gln Asp Asp Gly Pro Gln
40 45 50

cag gga cca ccc caa caa gga ggc cag cag caa caa ggt cca cca cct 293
Gln Gly Pro Pro Gln Gln Gly Gly Gln Gln Gln Gln Gly Pro Pro Pro
55 60 65 70

cct cag gga aag cca caa gga cca ccc caa cag gga ggc cag tct tgt 341
Pro Gln Gly Lys Pro Gln Gly Pro Pro Gln Gln Gly Gly Gln Ser Cys
75 80 85

tgc tgt gac aag taacagcctt ttttttaaatt tgtttttttat acaaaaaaaaaa 393
Cys Cys Asp Lys
90

aaaaaaaaa 401

<210> 10
<211> 961
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 45..467

<220>
<221> sig_peptide
<222> 45..116

<223> Von Heijne matrix
score 10.983328612314
seq LAGLCCLVPVSLA/ED

<220>

<221> polyA_site

<222> 946..961

<220>

<221> polyA_signal

<222> 927..932

<400> 10

```
ctcagcttca ggcaccacca ctgacctggg acagtgaatc gaca atg ccg tct tct      56
                                     Met Pro Ser Ser
gtc tcg tgg ggc atc ctc ctg ctg gca ggc ctg tgc tgc ctg gtc cct      104
Val Ser Trp Gly Ile Leu Leu Leu Ala Gly Leu Cys Cys Leu Val Pro
-20                                     -15      -10      -5
gtc tcc ctg gct gag gat ccc cag gga gat gct gcc cag aag aca gat      152
Val Ser Leu Ala Glu Asp Pro Gln Gly Asp Ala Ala Gln Lys Thr Asp
                                     1      5      10
aca tcc cac cat gat cag gat cac cca acc ttc aac aag atc acc ccc      200
Thr Ser His His Asp Gln Asp His Pro Thr Phe Asn Lys Ile Thr Pro
                                     15      20      25
aac ctg gct gag ttc gcc ttc agc cta tac cgc cag ctg gca cac cag      248
Asn Leu Ala Glu Phe Ala Phe Ser Leu Tyr Arg Gln Leu Ala His Gln
                                     30      35      40
tcc aac agc acc aat atc ttc ttc tcc cca gtg agc atc gct aca gcc      296
Ser Asn Ser Thr Asn Ile Phe Phe Ser Pro Val Ser Ile Ala Thr Ala
45      50      55      60
ttt gca atg ctc tcc ctg ggg acc aag gct gac act cac gat gaa atc      344
Phe Ala Met Leu Ser Leu Gly Thr Lys Ala Asp Thr His Asp Glu Ile
                                     65      70      75
ctg gag ggc ctg aat ttc aac ctc acg gag att ccg gag gct cag atc      392
Leu Glu Gly Leu Asn Phe Asn Leu Thr Glu Ile Pro Glu Ala Gln Ile
                                     80      85      90
cat gaa ggc ttc cag gaa ctc ctc cgt acc ctc aac cag cca gac agc      440
His Glu Gly Phe Gln Glu Leu Leu Arg Thr Leu Asn Gln Pro Asp Ser
95      100      105
cag ctc cag ctg acc acc ggc aag aat taatgctttc ttaactaagt      487
Gln Leu Gln Leu Thr Thr Gly Lys Asn
110      115
tactttgagt attaacaagt agcttatact gaggtgtgct tcaagggagc attatagggc      547
agaaaacaaa ttaaaatgaa atgttggtgt tatagccatg cattgaccat taaatgcatt      607
ttaatatctc ggcaaaatga gatttgaaac tcattagttt aactcttaaa acatagtgtt      667
gttaaacagt attttttggg ggcgggtgga tagcttttts ttctatcttg tacagaaaac      727
agtaatggaa tcaactcatgt tcagaacctt gtaaatatgt ctgtaatatt aatgcttccc      787
tttattaaca caaatataca ttcaaaaagg gagattttat ttagacacac attttctttt      847
caccctaatac tatttagatt catacttatt gagacgggaa aaactcgctg taaaataatg      907
ccaacctaga taatgctata ataaattatt ttgtacacaa aaaaaaaaaa aaaa      961
```

<210> 11

<211> 643

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 91..462

<220>

<221> sig_peptide

<222> 91..180

```

<223> Von Heijne matrix
      score 9.99993961154359
      seq LLFGFTLVSGTGA/EK

<220>
<221> polyA_site
<222> 628..643

<220>
<221> polyA_signal
<222> 607..612

<400> 11
accctaccc cacgccccct cccgcgcgcg cgggttaaata cccgcacctg agcatcggct      60
cacacctgca ccccgcccgg gcatagcacc atg cct gct tgt cgc cta ggc ccg      114
                               Met Pro Ala Cys Arg Leu Gly Pro
                               -30                               -25

cta gcc gcc gcc ctc ctc ctc agc ctg ctg ttc ggc ttc acc cta      162
Leu Ala Ala Ala Leu Leu Leu Ser Leu Leu Leu Phe Gly Phe Thr Leu
      -20                               -15                               -10

gtc tca ggc aca gga gca gag aag act ggc gtg tgc ccc gag ctc cag      210
Val Ser Gly Thr Gly Ala Glu Lys Thr Gly Val Cys Pro Glu Leu Gln
      -5                               1                               5                               10

gct gac cag aac tgc acg caa gag tgc gtc tcg gac agc gaa tgc gcc      258
Ala Asp Gln Asn Cys Thr Gln Glu Cys Val Ser Asp Ser Glu Cys Ala
      15                               20                               25

gac aac ctc aag tgc tgc agc gcg ggc tgt gcc acc ttc tgc tct ctg      306
Asp Asn Leu Lys Cys Cys Ser Ala Gly Cys Ala Thr Phe Cys Ser Leu
      30                               35                               40

ccc aat gat aag gag ggt tcc tgc ccc cag gtg aac att aac ttt ccc      354
Pro Asn Asp Lys Glu Gly Ser Cys Pro Gln Val Asn Ile Asn Phe Pro
      45                               50                               55

cag ctc ggc ctc tgt cgg gac cag tgc cag gtg gac agc cag tgt cct      402
Gln Leu Gly Leu Cys Arg Asp Gln Cys Gln Val Asp Ser Gln Cys Pro
      60                               65                               70

ggc cag atg aaa tgc tgc cgc aat ggc tgt ggg aag gtg tcc tgt gtc      450
Gly Gln Met Lys Cys Cys Arg Asn Gly Cys Gly Lys Val Ser Cys Val
      75                               80                               85                               90

act ccc aat ttc tgagctccag ccaccaccag gctgagcagt gaggagagaa      502
Thr Pro Asn Phe
agtttctgcc tggccctgca tctggttcca gccacactgc cctcccccttt ttcgggactc      562
tgtattccct cttgggctga ccacagcttc tccctttccc aaccaataaa gtaaccactt      622
tcagcaaaaa aaaaaaaaaa a
      643

<210> 12
<211> 577
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 28..399

<220>
<221> sig_peptide
<222> 28..117
<223> Von Heijne matrix
      score 9.99993961154359
      seq LLFGFTLVSGTGA/EK

<220>
<221> polyA_site
<222> 562..577

```

<220>

<221> polyA_signal

<222> 540..545

<400> 12

```

acctgcaccc cgcccgggca tagcacc atg cct gct tgt cgc cta ggc ccg cta      54
                        Met Pro Ala Cys Arg Leu Gly Pro Leu
                        -30                                -25
gcc gcc gcc ctc ctc ctc agc ctg ctg ctg ttc ggc ttc acc cta gtc      102
Ala Ala Ala Leu Leu Leu Ser Leu Leu Leu Phe Gly Phe Thr Leu Val
-20                                -15                                -10
tca ggc aca gga gca gag aag act ggc gtg tgc ccc gag ctc cag gct      150
Ser Gly Thr Gly Ala Glu Lys Thr Gly Val Cys Pro Glu Leu Gln Ala
-5                                1                                5                                10
gac cag aac tgc acg caa gag tgc gtc tcg gac agc gaa tgc gcc gac      198
Asp Gln Asn Cys Thr Gln Glu Cys Val Ser Asp Ser Glu Cys Ala Asp
15                                20                                25
aac atc aag tgc tgc agc gcg ggc tgt gcc acc ttc tgc tct ctg ccc      246
Asn Ile Lys Cys Cys Ser Ala Gly Cys Ala Thr Phe Cys Ser Leu Pro
30                                35                                40
aat gat aag gag ggt tcc tgc ccc cag gtg aac att aac ttt ccc cag      294
Asn Asp Lys Glu Gly Ser Cys Pro Gln Val Asn Ile Asn Phe Pro Gln
45                                50                                55
ctc ggc ctc tgt cgg gac cag tgc cag gtg gac agc cag tgt cct ggc      342
Leu Gly Leu Cys Arg Asp Gln Cys Gln Val Asp Ser Gln Cys Pro Gly
60                                65                                70                                75
cag atg aaa tgc tgc cgc aat ggc tgt ggg aag gtg tcc tgt gtc act      390
Gln Met Lys Cys Cys Arg Asn Gly Cys Gly Lys Val Ser Cys Val Thr
80                                85                                90
ccc aat ttc tgagctccag ccacaggctg agcagtgagg agagaaagtt      439
Pro Asn Phe
tctgcctggc cctgcattctg gttccagccc acctgccctc cccttttttcg ggactctgta      499
ttccctcttg ggctgaccac agctttctccc tttcccaacc aataaagtaa ccactttcag      559
ctaaaaaaaa aaaaaaaaaa      577

```

<210> 13

<211> 785

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 10..216

<220>

<221> sig_peptide

<222> 10..60

```

<223> Von Heijne matrix
      score 9.5148977432302
      FVFLLFVISLAAA/AH

```

<400> 13

```

aaccacatt atg ttt cct ggc ttt gtt ttt ttg ctc ttt gta ata agt ctt      51
                        Met Phe Pro Gly Phe Val Phe Leu Leu Phe Val Ile Ser Leu
                        -15                                -10                                -5
gct gct gct gct cat ctt tgg gtc ctg gcc gcc ttt atg ggc cgt atc      99
Ala Ala Ala Ala His Leu Trp Val Leu Ala Ala Phe Met Gly Arg Ile
1                                5                                10
acc gtg aag gtc tgc agc ttc act cct gag gcc agc aag acc gtg agc      147
Thr Val Lys Val Cys Ser Phe Thr Pro Glu Ala Ser Lys Thr Val Ser
15                                20                                25
cca cca gaa gga gcg aac aac tcg aga cgc act gct ttt aag agc tgt      195

```

Pro Pro Glu Gly Ala Asn Asn Ser Arg Arg Thr Ala Phe Lys Ser Cys
 30 35 40 45
 aac act cac cac aaa ggt ctg tagcatcact cctgaagtca gcgggaccac 246
 Asn Thr His His Lys Gly Leu
 50
 gaacctacca gaaagaagaa actccataca gatctgaaaa tctgaaggaa caaactccgg 306
 acacaccatc ttttaagaact gtaacactca ctgcgagggt ccacggcttc atttttgaag 366
 tcagccagac caagaaccca ccaattccgc acacacttct gctgtttcgt tccattctca 426
 aggaaaaact gtgatgggtc catttttaat cacatgcaaa ttcattgggca aatctttgct 486
 atggttatgt actttgattt agtagagttg tatcacatgt tcaactcatgt gccacaata 546
 gcagggtata ccaataggct gcttcaaaaac aatcatgaaa agtaggaggt ggtgtttcca 606
 aaggtagtgg tcttgccaga aaaaaatgga aatactgtga agataaaaaat aaacatctat 666
 taacacaaaa atagaaaaat aatataggta cacatacaca tactttttcta catatttttt 726
 gcagtcatat cagaggaaca atacagagga acaaaagtgt aaatcaaaaa aaaaaaaaaa 785

<210> 14
 <211> 1154
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 88..318

<220>
 <221> sig_peptide
 <222> 88..186
 <223> Von Heijne matrix
 score 9.33181480765342
 seq MVLLSFALTLCSA/FW

<220>
 <221> polyA_site
 <222> 1139..1154

<400> 14
 ctgtgggtgc ccaggaaggg actacacctc ttcgcagtgt tttatacctt tggtaatatc 60
 gcatcaattg ggagtaccat cttcctc atg gga cca gtg aaa cag ctg aag cga 114
 Met Gly Pro Val Lys Gln Leu Lys Arg
 -30 -25
 atg ttt gag cct act cgt ttg att gca act atc atg gtg ctg ttg agt 162
 Met Phe Glu Pro Thr Arg Leu Ile Ala Thr Ile Met Val Leu Leu Ser
 -20 -15 -10
 ttt gca ctt acc ctg tgt tct gcc ttt tgg tgg cat aac awg gga ctt 210
 Phe Ala Leu Thr Leu Cys Ser Ala Phe Trp Trp His Asn Xaa Gly Leu
 -5 1 5
 gca ctt atc ttc tgc att ttg cag tct ttg gca ttg acg tgg tac agc 258
 Ala Leu Ile Phe Cys Ile Leu Gln Ser Leu Ala Leu Thr Trp Tyr Ser
 10 15 20
 ctt tcc ttc ata cca ttt gca agg gat gct gtg aag aag tgt ttt gcc 306
 Leu Ser Phe Ile Pro Phe Ala Arg Asp Ala Val Lys Lys Cys Phe Ala
 25 30 35 40
 gtg tgt ctt gca taattcatgg ccagttttat gaagctttgg aaggcactat 358
 Val Cys Leu Ala
 ggacagaagc tgggtggacag ttttgtaact atcttcgaaa cctctgtctt acagacatgt 418
 gccttttatc ttgcagcaat gtgttgcttg tgattcgaaac atttgagggt tacttttgga 478
 agcaacaata cattctcgaa cctgaatgtc agtagcacag gatgagaagt gggttctgta 538
 tcttggtggag tggaatcttc ctcatgtacc tgtttcctct ctggatgttg tcccactgaa 598
 ttcccatgaa tacaaccta ttcagcaaca gcacataagc cttgggtgca agtgattccc 658
 aggtggcaaa aggcagcccc atcagagatc acgggagcaa cagtaaggga cagagttttg 718
 ggggtccactt gtccctcagc atggaagcca tcaccgtggg cctgcataga gtgagtctac 778
 ttctactctg gcatctgaga acaagtgact ctgctttaga caagcccctg gagagcctgg 838
 ccattggagtg aggtagaaaa gaagcacttt ttggtgggtat atgctgtttc tgaatttgaa 898

tctaagcatt	cttgggtgttc	tgattctgcg	cctgactgtg	gaatgtat	tctgacatta	958
ggtgcagact	ctttttctct	tgggttttat	ttcccatccc	tccctccctc	tccctgcata	1018
tctttcttta	tgacctcctg	gcactagatg	gggaaagaac	agagcagcag	cagcagaaaa	1078
tgctgctgtt	ttttttggac	gagagcccat	aggagttgaa	aaatcctgct	gctctcgtcc	1138
aaaaaaaaaa	aaaaaa					1154

<210> 15
 <211> 766
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 84..521

<220>
 <221> sig_peptide
 <222> 84..161
 <223> Von Heijne matrix
 score 8.99656184291183
 seq GLLWAFCAPGARA/EE

<220>
 <221> polyA_site
 <222> 751..766

<220>
 <221> polyA_signal
 <222> 728..733

<400> 15
 gaggaagagc gtttttgggga cgggggctgg tgaggctcac gttggagggc ttcgcgtctg 60
 cttcggagac cgtaaggata ttg atg acc atg aga tcc ctg ctc aga acc ccc 113
 Met Thr Met Arg Ser Leu Leu Arg Thr Pro
 -25 -20

ttc	ctg	tgt	ggc	ctg	ctc	tgg	gcc	ttt	tgt	gcc	cca	ggc	gcc	agg	gct	161
Phe	Leu	Cys	Gly	Leu	Leu	Trp	Ala	Phe	Cys	Ala	Pro	Gly	Ala	Arg	Ala	
-15						-10					-5					

gag gag cct gca gcc agc ttc tcc caa ccc ggc agc atg ggc ctg gat 209
 Glu Glu Pro Ala Ala Ser Phe Ser Gln Pro Gly Ser Met Gly Leu Asp
 1 5 10 15

aag aac aca gtg cac gac caa gag cat atc atg gag cat cta gaa ggt 257
 Lys Asn Thr Val His Asp Gln Glu His Ile Met Glu His Leu Glu Gly
 20 25 30

gtc atc aac aaa cca gag gcg gag atg tcg cca caa gaa ttg cag ctc 305
 Val Ile Asn Lys Pro Glu Ala Glu Met Ser Pro Gln Glu Leu Gln Leu
 35 40 45

cat tac ttc aaa atg cat gat tat gat ggc aat aat ttg ctt gat ggc 353
 His Tyr Phe Lys Met His Asp Tyr Asp Gly Asn Asn Leu Leu Asp Gly
 50 55 60

tta gaa ctc tcc aca gcc atc act cat gtc cat aag gag gaa ggg agt 401
 Leu Glu Leu Ser Thr Ala Ile Thr His Val Lys Glu Glu Gly Ser
 65 70 75 80

gaa cag gca cca cta atg agt gaa gat gaa ctg att aac ata ata gat 449
 Glu Gln Ala Pro Leu Met Ser Glu Asp Glu Leu Ile Asn Ile Ile Asp
 85 90 95

ggt gtt ttg aga gat gat gac aag aac aat gat gga tac att gac tat 497
 Gly Val Leu Arg Asp Asp Asp Lys Asn Asn Asp Gly Tyr Ile Asp Tyr
 100 105 110

gct gaa ttt gca aaa tca ctg cag tagatgttat ttggccatct cctgggttata 551
 Ala Glu Phe Ala Lys Ser Leu Gln
 115 120

tacaaatgtg acccgtgata atgtgattga acacttttagt aatgcgaaat aactcatttc 611

14

caactactgc	tgcagcattt	tggtaaaaac	ctgtagcgat	tcgttacact	ggggtgagaa	671
gagataagag	aaatgaaaga	gaagagaaat	gggacatcta	atagtcoccta	agtgstatta	731
aataccttat	tggaacaagca	aaaaaaaaaa	aaaaa			766

<210> 16
 <211> 808
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 155..340

<220>
 <221> sig_peptide
 <222> 155..292
 <223> Von Heijne matrix
 score 8.64329745298384
 AVLLLILFAIVFG/LL

<220>
 <221> polyA_site
 <222> 754..808

<220>
 <221> polyA_signal
 <222> 728..733

<400> 16	
cttttctctt	caacagttgc
cttcagctca	gccctccaca
gaccagaac	agctaccagc
	agaatcagat
	tctc atg gac caa ctg gta ttc aaa
	Met Asp Gln Leu Val Phe Lys
	-45 -40
gag aca atc tgg aat gat gcg ttc tgg cag aac ccc tgg gac cag ggg	223
Glu Thr Ile Trp Asn Asp Ala Phe Trp Gln Asn Pro Trp Asp Gln Gly	
	-35 -30 -25
ggc ctg gca gtg att atc tta ttc atc acc gct gtc ctg ctt ctc atc	271
Gly Leu Ala Val Ile Ile Leu Phe Ile Thr Ala Val Leu Leu Leu Ile	
	-20 -15 -10
tta ttt gcc atc gtg ttt ggt tta ctc act tcc aca gaa aac act cag	319
Leu Phe Ala Ile Val Phe Gly Leu Leu Thr Ser Thr Glu Asn Thr Gln	
	-5 1 5
tgt gaa gcg ggt gaa gag gag tgacctgact tgctggggac tgagatggca	370
Cys Glu Ala Gly Glu Glu Glu	
10 15	
gcaggggagg	cgagctgacc
ggctcagggg	ccaagccctg
tgtatctggc	ttaggggttg
tggtccttg	accctattgt
aggatggctc	caactgcctca
cacaggccca	gccagcaatt
aaaaattcct	tttcaaattc
aaaaaaaaaa	aaaaaaaaaa
	430 490 550 610 670 730 790 808

<210> 17
 <211> 772
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 155..340

```

<220>
<221> sig_peptide
<222> 155..292
<223> Von Heijne matrix
      score 8.64329745298384
      AVLLLILFAIVFG/LL

<220>
<221> polyA_site
<222> 751..772

<220>
<221> polyA_signal
<222> 728..733

<400> 17
cttttctct caacagttgc ttctttgagt caggggtgcag ctctgggtcac ctggcggcct      60
cttcagctca gccctccaca aagtgtgagc ctgaaggacc accctgaatt gcccttgtag      120
gaccagaac agcaaccagc agaatcagat tctc atg gac caa ttg gta ttc aaa      175
                               Met Asp Gln Leu Val Phe Lys
                               -45                               -40

gag aca atc tgg aat gat gcg ttc tgg cag aac ccc tgg gac cag ggg      223
Glu Thr Ile Trp Asn Asp Ala Phe Trp Gln Asn Pro Trp Asp Gln Gly
                               -35                               -30                               -25

ggc ctg gca gtg att atc tta ttc atc acc gct gtc ctg ctt ctc atc      271
Gly Leu Ala Val Ile Ile Leu Phe Ile Thr Ala Val Leu Leu Leu Ile
                               -20                               -15                               -10

tta ttt gcc atc gtg ttt ggt tta ctc act tcc aca gaa aac act cag      319
Leu Phe Ala Ile Val Phe Gly Leu Leu Thr Ser Thr Glu Asn Thr Gln
                               -5                               1                               5

tgt gaa gcg ggt gaa gag gag tgacctgact tgctggggac tgagatggca      370
Cys Glu Ala Gly Glu Glu
10                               15

gcaggggagg cgagctgacc tgccccatt ccagtgggtgg gcccttcgc ggttccctct      430
ggctcagggg ccaagccctg gtgtcttctt tttccaccag gaaaaagtct agtaaaatac      490
tgtatctggc ttagggttgg tcagactagt aagatgggga ggctgggtctg agaccaattc      550
tggtccttgg accctattgt ttttaggggt ccccgaccag aaccctaaaa gcacatggag      610
aggatggctc cactgcctca ggtggaagga gctatggcta acaaggttct ctaacaggct      670
cacaggcca gccagcaatt tcacaaatcc ttgacagaga aagacacaac caaatgaaat      730
aaaaattcct tttcaaatct aaaaaaaaaa aaaaatgcga aa      772

<210> 18
<211> 546
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 13..498

<220>
<221> sig_peptide
<222> 13..153
<223> Von Heijne matrix
      score 4.65474235719409
      seq LALSSLRSLLLFA/GM

<220>
<221> polyA_site
<222> 533..546

<220>

```


<221> polyA_signal

<222> 514..519

<400> 18

```

aagcgctgac gc atg cgc ata gct aac cgc acc cgg ttc agc tcg cct ttc      51
               Met Arg Ile Ala Asn Arg Thr Arg Phe Ser Ser Pro Phe
               -45               -40               -35
ttg gcc aga ggc gcc ggt tgg act cac ggg cgg ggc atg atg gtg gtg      99
Leu Ala Arg Gly Ala Gly Trp Thr His Gly Arg Gly Met Met Val Val
               -30               -25               -20
ggc acg ggc acc tcg ctg gcg ctc tcc tcc ctc cgg tcc ctg ctg ctc      147
Gly Thr Gly Thr Ser Leu Ala Leu Ser Ser Leu Arg Ser Leu Leu Leu
               -15               -10               -5
ttt gct ggg atg cag atg tac agc cgt cag ctg gcc tcc acc gag tgg      195
Phe Ala Gly Met Gln Met Tyr Ser Arg Gln Leu Ala Ser Thr Glu Trp
               1               5               10
ctc acc atc cag ggc ggc ctg ctt ggt tcg ggt ctc ttc gtg ttc tcg      243
Leu Thr Ile Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe Val Phe Ser
15               20               25               30
ctc act gcc ttc aat aat ctg gag aat ctt gtc ttt ggc aaa gga ttc      291
Leu Thr Ala Phe Asn Asn Leu Glu Asn Leu Val Phe Gly Lys Gly Phe
               35               40               45
caa gca aag atc ttc cct gag att ctc ctg tgc ctc ctg ttg gct ctc      339
Gln Ala Lys Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu Leu Ala Leu
               50               55               60
ttt gca tct ggc ctc atc cac cga gtc tgt gtc acc acy tgc ttc atc      387
Phe Ala Ser Gly Leu Ile His Arg Val Cys Val Thr Thr Cys Phe Ile
               65               70               75
ttc tcc atg gtt ggt ctg tac tac atc aac aag atc tcc tcc acc ctg      435
Phe Ser Met Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser Ser Thr Leu
               80               85               90
tac cag gca gca gct cca gtc ctc aca cca gcc aag gtc aca ggc aag      483
Tyr Gln Ala Ala Ala Pro Val Leu Thr Pro Ala Lys Val Thr Gly Lys
95               100               105               110
agc aag aag aga aac tgaccctgaa tgttcaataa agttgattct ttgcaaaaaa      538
Ser Lys Lys Arg Asn
               115
aaaaaaaaa      546

```

<210> 19

<211> 767

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 132..401

<220>

<221> sig_peptide

<222> 132..236

<223> Von Heijne matrix

score 8.35677668307059

seq LGLGLVTLTASWA/AL

<220>

<221> polyA_site

<222> 752..767

<220>

<221> polyA_signal

<222> 741..746

```

<400> 19
ccttcagcgc ccggccgaga gcccggtggc tgcgactgag agcctgggta ctgccggcac      60
ctttggatcc tcggccaatt ctctgcttcg cctacacttg gcgtgcggtg taccagcctc      120
cccgcgtcac c atg gaa aca gga gcc tct gca tcc atc cca gag ctg atc      170
          Met Glu Thr Gly Ala Ser Ala Ser Ile Pro Glu Leu Ile
          -35          -30          -25

tgt gaa gct atg aga aga atc tgg agc ctg gga ttg ggg ttg gtg act      218
Cys Glu Ala Met Arg Arg Ile Trp Ser Leu Gly Leu Gly Leu Val Thr
          -20          -15          -10

ctg acg gcc agc tgg gca gct ctt ttc cac gat ggc ttt gcg gtt ctt      266
Leu Thr Ala Ser Trp Ala Ala Leu Phe His Asp Gly Phe Ala Val Leu
          -5          1          5          10

gga gga aac att gtg agc gat ctc agc aca gta aga ttt gtt gca caa      314
Gly Gly Asn Ile Val Ser Asp Leu Ser Thr Val Arg Phe Val Ala Gln
          15          20          25

cag cag cac ttc cag ctc ctt gac gtg tgg acc agg aat ttc cgg aag      362
Gln Gln His Phe Gln Leu Leu Asp Val Trp Thr Arg Asn Phe Arg Lys
          30          35          40

cca ctg ggc agc atg tgc ttt gtt ttc ttg ttg ctc cca taatcaatgt      411
Pro Leu Gly Ser Met Cys Phe Val Phe Leu Leu Leu Pro
          45          50          55

tgggcatcaa gatctaacc ttgaaccttc cacgaaccct gttgtcaata cctctggggtt      471
tccgccgggtt acgcttaatt ttgacatatc ggtcagactg gtgccggata aacttcttgg      531
ttctcttttt gacgatcttg ggcttcccaa gggttctgag gacggtttat aaaagtgcac      591
tgaggggagg ctgagacagg agaattgctt gaaccgggga ggcggagggtt gcagtgagcc      651
gagatcgtgc ccacccctagg cgacagagag agactctatg tcaaaaataa ataaataaat      711
aaataaataa ataaataaat aaataaataa ataaaatgcc aaaaaaaaaa aaaaaa      767

```

```

<210> 20
<211> 596
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 41..478

```

```

<220>
<221> sig_peptide
<222> 41..115
<223> Von Heijne matrix
      score 8.18796433724836
      seq ATLLLVLCLQLGA/NK

```

```

<220>
<221> polyA_site
<222> 581..596

```

```

<220>
<221> polyA_signal
<222> 560..565

```

```

<400> 20
accacttctc tgggacacat tgccttctgt tttctccagc atg cgc ttg ctc cag      55
          Met Arg Leu Leu Gln
          -25

ctc ctg ttc agg gcc agc cct gcc acc ctg ctc ctg gtt ctc tgc ctg      103
Leu Leu Phe Arg Ala Ser Pro Ala Thr Leu Leu Leu Val Leu Cys Leu
-20          -15          -10          -5

cag ttg ggg gcc aac aaa gct cag gac aac act cgg aag atc ata ata      151
Gln Leu Gly Ala Asn Lys Ala Gln Asp Asn Thr Arg Lys Ile Ile Ile
          1          5          10

aag aat ttt gac att ccc aag tca gta cgt cca aat gac gaa gtc act      199

```

Lys	Asn	Phe	Asp	Ile	Pro	Lys	Ser	Val	Arg	Pro	Asn	Asp	Glu	Val	Thr		
	15						20				25						
gca	gtg	ctt	gca	gtt	caa	aca	gaa	ttg	aaa	gaa	tgc	atg	gtg	gtt	aaa	247	
Ala	Val	Leu	Ala	Val	Gln	Thr	Glu	Leu	Lys	Glu	Cys	Met	Val	Val	Lys		
	30						35				40						
act	tac	ctc	att	agc	agc	atc	cct	cta	caa	ggk	gcm	ttt	aac	tat	arg	295	
Thr	Tyr	Leu	Ile	Ser	Ser	Ile	Pro	Leu	Gln	Gly	Ala	Phe	Asn	Tyr	Xaa		
	45				50					55				60			
tat	act	gcc	tgc	cta	tgt	gac	gac	aat	cca	aaa	acc	ttc	tac	tgg	gac	343	
Tyr	Thr	Ala	Cys	Leu	Cys	Asp	Asp	Asn	Pro	Lys	Thr	Phe	Tyr	Trp	Asp		
				65				70						75			
ttt	tac	acc	aac	aga	act	gtg	caa	att	gca	gcc	gtc	gtt	gat	gtt	att	391	
Phe	Tyr	Thr	Asn	Arg	Thr	Val	Gln	Ile	Ala	Ala	Val	Val	Asp	Val	Ile		
			80				85						90				
cgg	gaa	tta	ggc	atc	tgc	cct	gat	gat	gct	gct	gta	atc	ccc	atc	aaa	439	
Arg	Glu	Leu	Gly	Ile	Cys	Pro	Asp	Asp	Ala	Ala	Val	Ile	Pro	Ile	Lys		
		95				100					105						
aac	aac	cgg	ttt	tat	act	att	gaa	atc	cta	aag	gta	gaa	taatggaagc			488	
Asn	Asn	Arg	Phe	Tyr	Thr	Ile	Glu	Ile	Leu	Lys	Val	Glu					
	110					115					120						
cctgtctgtt	tgccacaccc	aggtgatttc	ctctaaagaa	acttggctgg	aatttctgct											548	
gtggtctata	aaataaactt	cttaacatgc	tcaaaaaaaaa	aaaaaaaaa												596	

<210> 21

<211> 672

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 7..366

<220>

<221> sig_peptide

<222> 7..102

<223> Von Heijne matrix
 score 7.64815044221151
 VTLGIGFFALASA/LW

<220>

<221> polyA_site

<222> 641..672

<400> 21

agtc	cat	atg	gct	tca	cct	aaa	ggt	ttt	ttt	aat	tac	ttg	acc	tat	ttt	48	
		Met	Ala	Ser	Pro	Lys	Gly	Phe	Phe	Asn	Tyr	Leu	Thr	Tyr	Phe		
			-30					-25					-20				
ctt	gct	gct	ggt	gct	gtc	act	ttg	gga	att	ggt	ttc	ttt	gct	ttg	gca	96	
Leu	Ala	Ala	Gly	Ala	Val	Thr	Leu	Gly	Ile	Gly	Phe	Phe	Ala	Leu	Ala		
			-15					-10					-5				
tca	gct	ttg	tgg	ttc	ctg	att	tgc	aaa	cga	aga	gaa	ata	ttt	caa	aat	144	
Ser	Ala	Leu	Trp	Phe	Leu	Ile	Cys	Lys	Arg	Arg	Glu	Ile	Phe	Gln	Asn		
		1				5					10						
tcc	aaa	ttt	aaa	gca	att	gat	gag	aga	tgc	agg	caa	aga	cca	tca	atg	192	
Ser	Lys	Phe	Lys	Ala	Ile	Asp	Glu	Arg	Cys	Arg	Gln	Arg	Pro	Ser	Met		
	15				20				25					30			
gcg	aaa	att	aaa	tct	cat	tct	cag	tgt	gtt	ttt	att	tct	cga	aat	ttt	240	
Ala	Lys	Ile	Lys	Ser	His	Ser	Gln	Cys	Val	Phe	Ile	Ser	Arg	Asn	Phe		
				35				40					45				
cat	act	ggg	aga	ttc	caa	tta	cag	caa	tta	aag	atc	att	cta	aag	atg	288	
His	Thr	Gly	Arg	Phe	Gln	Leu	Gln	Leu	Lys	Ile	Ile	Leu	Lys	Met			
			50					55				60					
aac	ccc	aac	ttg	caa	caa	aaa	ata	tca	ttt	gtg	atc	cct	cag	aga	cca	336	

Asn Pro Asn Leu Gln Gln Lys Ile Ser Phe Val Ile Pro Gln Arg Pro
 65 70 75
 gct cca caa caa atc gca gca gtg tta cat taagcttatac aacattacca 386
 Ala Pro Gln Gln Ile Ala Ala Val Leu His
 80 85
 tctgattctt attacagcca aagtatagaa gcagctgatg actgggttttc tgatgattct 446
 ctagtgaataa ggaactctcc aatgccttct ctcggggaac ctctaataaga aaaagtattt 506
 tcatacctgt caaccatttc attagaagag ggtactgaaa gtgtactgaa tgacacttta 566
 tgaccatcaa aaagatgact acattaaggg aaaatgttca tgaagaaaca cagaggttga 626
 aatataaaac cttcaacaga aaaaaaaaaa aaaatgcgaa aagctt 672

<210> 22
 <211> 438
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 84..317

<220>
 <221> sig_peptide
 <222> 84..140
 <223> Von Heijne matrix
 score 7.64030745849671
 seq ALVLALMISMISA/DS

<220>
 <221> polyA_site
 <222> 421..438

<220>
 <221> polyA_signal
 <222> 397..402

<400> 22
 atagaaaagg acatctcttg agacttcact tcagcttcac tgacttcttg actctcctct 60
 tgagtaaaag gactcagcca act atg aag ttt ttt gtc ttt gct tta gtc ttg 113
 Met Lys Phe Phe Val Phe Ala Leu Val Leu
 -15 -10
 gct ctc atg att tcc atg att agc gct gat tca cat gaa aag aga cat 161
 Ala Leu Met Ile Ser Met Ile Ser Ala Asp Ser His Glu Lys Arg His
 -5 1 5
 cat ggg tat aga aga aaa ttc cat gaa aag cat cat tca tac cat atc 209
 His Gly Tyr Arg Arg Lys Phe His Glu Lys His His Ser Tyr His Ile
 10 15 20
 aca cta cta cca ctt ttt gaa gaa tca tca aag agc aat gca aat gaa 257
 Thr Leu Leu Pro Leu Phe Glu Glu Ser Ser Lys Ser Asn Ala Asn Glu
 25 30 35
 aaa cac tat aat tta ctg tat act ctt tgt ttc agg ata ctt gcc ttt 305
 Lys His Tyr Asn Leu Leu Tyr Thr Leu Cys Phe Arg Ile Leu Ala Phe
 40 45 50 55
 tca att gtc act tgatgatata attgcaattt aaactgttaa gctgtgttca 357
 Ser Ile Val Thr
 gtactgtttc tgaataatag aaatcacttc tctaaaagca ataaatttca agcacatttt 417
 taaataaaaa aaawawaaaa a 438

<210> 23
 <211> 617
 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS
<222> 141..401

<220>
<221> sig_peptide
<222> 141..200
<223> Von Heijne matrix
score 7.42708462116258
LLVFLAGFPVLDA/ND

<220>
<221> polyA_site
<222> 603..617

<400> 23
gatttctccc ggaacctctg ctcagcctgg tgaaccacac aggcccgagt ttcacccagt 60
ccccactcca cgggtgcagct gcggcttatt tctcagccca gcgagatgcc agccttctg 120
tcccgggcca gcgctctgac atg cag aag gtg acc ctg ggc ctg ctt gtg ttc 173
Met Gln Lys Val Thr Leu Gly Leu Leu Val Phe
-20 -15 -10
ctg gca ggc ttt cct gtc ctg gac gcc aat gac cta gaa gat aaa aac 221
Leu Ala Gly Phe Pro Val Leu Asp Ala Asn Asp Leu Glu Asp Lys Asn
-5 1 5
agt cct ttc tac tat gac tgg cac agc ctc cag gtt ggc ggg ctc atc 269
Ser Pro Phe Tyr Tyr Asp Trp His Ser Leu Gln Val Gly Gly Leu Ile
10 15 20
tgc gct ggg gtt ctg tgc gcc atg ggc atc atc atc gtc atg agt gca 317
Cys Ala Gly Val Leu Cys Ala Met Gly Ile Ile Ile Val Met Ser Ala
25 30 35
aaa tgc aaa tgc aag ttt ggc cag aag tcc ggt cac cat cca ggg gag 365
Lys Cys Lys Cys Lys Phe Gly Gln Lys Ser Gly His His Pro Gly Glu
40 45 50 55
act cca cct ctc atc acc cca ggc tca gcc caa agc tgatgaggac 411
Thr Pro Pro Leu Ile Thr Pro Gly Ser Ala Gln Ser
60 65
agaccagctg aaattgggtg gaggaccgtt ctctgtcccc aggtcctgtc tctgcacaga 471
aacttgaact ccaggatgga attcttcctc ctctgctggg actcctttgc atggcagggc 531
ctcatctcac ctctcgcaag aggtctcttt tgttcaattt ttttttatct aaaatgattg 591
tgcctctgcc caaaaaaaaa aaaaaa 617

<210> 24
<211> 673
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 69..383

<220>
<221> sig_peptide
<222> 69..125
<223> Von Heijne matrix
score 7.20796835452081
VSIMLLLVTVSDC/AV

<400> 24
acaaggctga gcgggaggaa gcgagaggca tctaagcagg cagtgttttg ccttcacccc 60
aagtgacc atg aga ggt gcc acg cga gtc tca atc atg ctc ctc cta gta 110
Met Arg Gly Ala Thr Arg Val Ser Ile Met Leu Leu Leu Val
-15 -10
act gtg tct gac tgt gct gtg atc aca ggg gcc tgt gag cga gat gtc 158
Thr Val Ser Asp Cys Ala Val Ile Thr Gly Ala Cys Glu Arg Asp Val

21

```

-5          1          5          10
cag tgt ggg gca ggc acc tgc tgt gcc atc agc ctg tgg ctt cga ggg      206
Gln Cys Gly Ala Gly Thr Cys Cys Ala Ile Ser Leu Trp Leu Arg Gly
          15          20          25
ctg cgg atg tgc acc ccg ctg ggg cgg kaa ggc gag gag tgc cac ccc      254
Leu Arg Met Cys Thr Pro Leu Gly Arg Xaa Gly Glu Glu Cys His Pro
          30          35          40
ggc agc cac aag atc ccc ttc ttc agg aaa cgc aag cac cac acc tgt      302
Gly Ser His Lys Ile Pro Phe Arg Lys Arg Lys His His Thr Cys
          45          50          55
cct tgc ttg ccc aac ctg ctg tgc tcc agg ttc ccg gac ggc agg tac      350
Pro Cys Leu Pro Asn Leu Leu Cys Ser Arg Phe Pro Asp Gly Arg Tyr
          60          65          70          75
cgc tgc tcc atg gac ttg aag aac atc aat ttt taggcgcttg cctggtctca      403
Arg Cys Ser Met Asp Leu Lys Asn Ile Asn Phe
          80          85
ggatacccac catccttttc ctgagcacag cctggatttt tattttctgcc atgaaaccca      463
gctcccatga ctctcccagt ccctasactg actaccctga tctctcttgt ctagtacgca      523
catatgcaca caggcagaca tacctcccat catgacatgg tccccaggct ggcctgagga      583
tgtcacagct tgaggctgtg gtgtgaaagg tggccagcct gggttctcttc cctgctcagg      643
ctgcactcaa aaaaaaaaaa aaatgcgaaa      673

```

<210> 25
 <211> 678
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 69..383

<220>
 <221> sig_peptide
 <222> 69..125
 <223> Von Heijne matrix
 score 7.20796835452081
 VSIMLLLVTVSDC/AV

<220>
 <221> polyA_site
 <222> 652..678

```

<400> 25
acaaggctga gcgggaggaa gcgagaggca tctaagcagg cagtgttttg ccttcacccc      60
aagtgacc atg aga ggt gcc acg cga gtc tca atc atg ctc ctc cta gta      110
          Met Arg Gly Ala Thr Arg Val Ser Ile Met Leu Leu Leu Val
          -15          -10
act gtg tct gac tgt gct gtg atc aca ggg gcc tgt gag cga gat gtc      158
Thr Val Ser Asp Cys Ala Val Ile Thr Gly Ala Cys Glu Arg Asp Val
-5          1          5          10
cag tgt ggg gca ggc acc tgc tgt gcc atc agc ctg tgg ctt cga ggg      206
Gln Cys Gly Ala Gly Thr Cys Cys Ala Ile Ser Leu Trp Leu Arg Gly
          15          20          25
ctg cgg atg tgc acc ccg ctg ggg cgg gaa ggc gag gag tgc cac ccc      254
Leu Arg Met Cys Thr Pro Leu Gly Arg Glu Gly Glu Glu Cys His Pro
          30          35          40
ggc agc cac aag atc ccc ttc ttc agg aaa cgc aag cac cac acc tgt      302
Gly Ser His Lys Ile Pro Phe Phe Arg Lys Arg Lys His His Thr Cys
          45          50          55
cct tgc ttg ccc aac ctg ctg tgc tcc agg ttc ccg gac ggc agg tac      350
Pro Cys Leu Pro Asn Leu Leu Cys Ser Arg Phe Pro Asp Gly Arg Tyr
          60          65          70          75
cgc tgc tcc atg gac ttg aag aac atc aat ttt taggcgcttg cctggtctca      403

```

Arg Cys Ser Met Asp Leu Lys Asn Ile Asn Phe

80

85

ggataccac	catccttttc	ctgagcacag	cctggatttt	tattttctgcc	atgaaaccca	463
gctcccatga	ctctcccagt	ccctacactg	actaccctga	tctctcttgt	ctagtacgca	523
catatgcaca	caggcagaca	tacctcccat	catgacatgg	tccccaggct	ggcctgagga	583
tgtcacagct	tgaggctgtg	gtgtgaaagg	tggccagcct	ggttctcttc	cctgctcagg	643
ctgcactcaa	aaaaaaaaaa	aaatgcgaaa	agctt			678

<210> 26

<211> 860

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 129..434

<220>

<221> sig_peptide

<222> 129..236

<223> Von Heijne matrix

score 7.04760006506154

seq PFFLITLVGVVVA/VV

<220>

<221> polyA_site

<222> 845..860

<220>

<221> polyA_signal

<222> 826..831

<400> 26

tccggggccc	gggagccaac	cgagggcggtt	cctgtcgggg	ctgcagcggc	gggagggagc	60
ccagtggagg	cgccctcccg	aagcgccact	gcccatgctg	accacccagc	cctccggctg	120
ctgatgtc	atg agt aac acc act gtg	ccc aat gcc ccc cag gcc aac agc				170
	Met Ser Asn Thr Thr Val	Pro Asn Ala Pro Gln Ala Asn Ser				
	-35	-30	-25			
gac tcc atg gtg ggc tat gtg ttg ggg ccc ttc ttc ctc atc acc ctg						218
Asp Ser Met Val Gly Tyr Val Leu Gly Pro Phe Phe Leu Ile Thr Leu						
	-20	-15	-10			
gtc ggg gtg gtg gtg gct gtg gta atg tat gta cag aag aaa aag cgg						266
Val Gly Val Val Val Ala Val Val Met Tyr Val Gln Lys Lys Lys Arg						
	-5	1	5	10		
gtg gac cgg ctg cgc cat cac ctg ctc ccc atk tac agc tat gac cca						314
Val Asp Arg Leu Arg His His Leu Leu Pro Xaa Tyr Ser Tyr Asp Pro						
	15	20	25			
gct gag gaa ctg cat gag gct gag car cag ctg ctc tct gac atg gga						362
Ala Glu Glu Leu His Glu Ala Glu Gln Gln Leu Leu Ser Asp Met Gly						
	30	35	40			
gac ccc aag gtg gta cat ggc tgg cag akt ggc tac cag cac aag cgg						410
Asp Pro Lys Val Val His Gly Trp Gln Xaa Gly Tyr Gln His Lys Arg						
	45	50	55			
atg cca ctg ctg gat gtc aag acg tgacctgacc cccttgcccc acccttcaga						464
Met Pro Leu Leu Asp Val Lys Thr						
	60	65				
gcctgggggtc	ctggactgcc	tggggccctg	ccatctgctt	cccctgctgt	cacctggctc	524
cccctgctgg	gtgctgggtc	tccatttctc	cctccacca	ccctcagcag	catctgcttc	584
ccatgccctc	accatcacct	cactgcccc	aggccttctg	ccctttgtgg	gtgttgagct	644
caccgcccac	ccacaggcac	tcataggaag	aggctttcct	tctgggatgg	cggcggctgg	704
tagacacctt	tgttttctct	agccctcctg	ggctgggctt	gggcacaaat	ccccaggcag	764
gcttttgagt	tgtttccatg	gtgatggggc	cagatgtata	gtattcagta	tatatattgt	824
aaataaaatg	ttttgtggct	aaaaaaaaaa	aaaaaa			860

<210> 27
 <211> 1443
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 30..440

<220>
 <221> sig_peptide
 <222> 30..83
 <223> Von Heijne matrix
 score 7.06387130837154
 seq VLGLVLLSVTVQG/KV

<220>
 <221> polyA_site
 <222> 1428..1443

<220>
 <221> polyA_signal
 <222> 1409..1414

<400> 27
 cagcctagca ctctgaccta gcagtcac atg aag gct ctc att gtt ctg ggg 53
 Met Lys Ala Leu Ile Val Leu Gly
 -15
 ctt gtc ctc ctt tct gtt acg gtc cag ggc aag gtc ttt gaa agg tgt 101
 Leu Val Leu Leu Ser Val Thr Val Gln Gly Lys Val Phe Glu Arg Cys
 -10 -5 1 5
 gag ttg gcc aga act ctg aaa aga ttg gga atg gat ggc tac agg gga 149
 Glu Leu Ala Arg Thr Leu Lys Arg Leu Gly Met Asp Gly Tyr Arg Gly
 10 15 20
 atc agc cta gca aac tgg atg tgt ttg gcc aaa tgg gag agt ggt tac 197
 Ile Ser Leu Ala Asn Trp Met Cys Leu Ala Lys Trp Glu Ser Gly Tyr
 25 30 35
 aac aca cga gct aca aac tac aat gct gga gac aga agc act gat tat 245
 Asn Thr Arg Ala Thr Asn Tyr Asn Ala Gly Asp Arg Ser Thr Asp Tyr
 40 45 50
 ggg ata ttt cag atc aat agc cgc tac tgg tgt aat gat ggc aaa acc 293
 Gly Ile Phe Gln Ile Asn Ser Arg Tyr Trp Cys Asn Asp Gly Lys Thr
 55 60 65 70
 cca gga gca gtt aat gcc tgt cat tta tcc tgc agt ggg tgg cat gga 341
 Pro Gly Ala Val Asn Ala Cys His Leu Ser Cys Ser Gly Trp His Gly
 75 80 85
 gaa atc gtt gtc aaa aca gag atg tcc gtc agt atg ttc aag gtt gtg 389
 Glu Ile Val Val Lys Thr Glu Met Ser Val Ser Met Phe Lys Val Val
 90 95 100
 gag tgt aac tcc aga att ttc ctt ctt cag ctc att ttg tct ctc tca 437
 Glu Cys Asn Ser Arg Ile Phe Leu Leu Gln Leu Ile Leu Ser Leu Ser
 105 110 115
 cat taagggagta ggaattaagt gaaaggtcac actaccatta ttttcccctt 490
 His
 caaacaata atatTTTTTtac agaagcagga gcaaaatatg gccttttcttc taagagatat 550
 aatgttcact aatgtggtta ttttatatta agcctacaac attttttcagt ttgcaaatag 610
 aactaatact ggtgaaaatt tacctaaaac cttggttatc aaatacatct ccagtacatt 670
 ccgttctttt tttttttttt tttttttttt ttgagacagt ctgctctgt cgcccaggct 730
 ggagtgcagt ggcgcaatct cggctcactg caacctccac ctcccgggtt caccgcatc 790
 tctgcctca gcctcccgag tagctgggat tacgggcgcc cgccaccacg cccggctaata 850
 tttttgtatt tttagtagag acagggtttc accgttttag ccaggatggt ctcgatctcc 910
 tgaccttgtg atccaccac ctcggcctcc caaagtgtcg cgattacagg cgtgagccac 970


```

tgcgcccggc cacattcagt tyttatcaaa gaaataaccc agacttaatc ttgaatgata 1030
cgattatgcc caatattaag taaaaaatat aagaaaaggt katctttaat agatcttagg 1090
caaaatacca gctgatgaag gcatctgatg ccttcacatg ttcagtcac tccaaaaaca 1150
gtaaaaataa ccactttttg ttgggcaata tgaaattttt aaaggagtag aataccaaat 1210
gatagaaaca gactgcctga attgagaatt ttgatttttt aaagtgtgtt tctttctaaa 1270
ttgctgttcc ttaatttgat taatttaatt catgtattat gattaaatct gakgcakatg 1330
agcttacaag tattgaaata attactaatt aatcacaaat gtgaagttat gcatgatgta 1390
aaaaatacaa acattctaatt taaaggcttt gcaacacaaa aaaaaaaaaa aaa 1443

```

<210> 28
 <211> 622
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 82..384

<220>
 <221> sig_peptide
 <222> 82..168
 <223> Von Heijne matrix
 score 6.9066614461057
 seq VCMCVFSTQGALG/EM

<220>
 <221> polyA_site
 <222> 607..622

<220>
 <221> polyA_signal
 <222> 570..575

```

<400> 28
gctacctcct ggtgtctgtt gcctggtgtg tgggaaaccc ccacacgctg gaccacagaa 60
gtcttctgtg ttgactgttg a atg agg gaa gag aaa aag ccc ttt gag aga 111
                               Met Arg Glu Glu Lys Lys Pro Phe Glu Arg
                               -25                               -20

gag aga gag agt gtg tgt gtg tgt atg tgt gtg ttt tcc act caa gga 159
Glu Arg Glu Ser Val Cys Val Cys Met Cys Val Phe Ser Thr Gln Gly
                               -15                               -10                               -5

gct ttg ggg gaa atg gct gca cac ttc ata gat gaa aag ctg agg ccc 207
Ala Leu Gly Glu Met Ala Ala His Phe Ile Asp Glu Lys Leu Arg Pro
                               1                               5                               10

agt gag ggg aat ggt cac aga ggg aca ctg gat agc ctg agt tcg gac 255
Ser Glu Gly Asn Gly His Arg Gly Thr Leu Asp Ser Leu Ser Ser Asp
                               15                               20                               25

caa gag tcc tac atc ccc tcc acc gcc gac ccc acc cag gct ggc cct 303
Gln Glu Ser Tyr Ile Pro Ser Thr Ala Asp Pro Thr Gln Ala Gly Pro
30                               35                               40                               45

gag ctg ctg cac aag aac ttg ccc gtg acc tcc cga tcc cag ccg ctg 351
Glu Leu Leu His Lys Asn Leu Pro Val Thr Ser Arg Ser Gln Pro Leu
                               50                               55                               60

ccc tct gac ctc gcg atc cca gcc gct gcc ctc tgacctcccg atcccagccg 404
Pro Ser Asp Leu Ala Ile Pro Ala Ala Leu
                               65                               70

ctgccctcgc tgagcttgga ccacgtgcct gagaccgctc tttcctcatt taaaccctgc 464
tcggggtggg gggagctcca caccggcggg gaacgtgctt gttcacagtg tgttcttagc 524
ctctagaaca gagcctcagg cacaggaggc acttcacatt tattgaataa atagatgaat 584
gaatcagtgt catcctccca acaaaaaaaaa aaaaaaaaaa 622

```

<210> 29
 <211> 658

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 114..497

<220>
<221> sig_peptide
<222> 114..191
<223> Von Heijne matrix
score 6.8337397741578
seq VFLSLFLIQLLIS/FS

<400> 29
ctcagcaccc agggcggtgg taggtcacag tctctgggcg ggtctcagtg tocaacactg 60
tagctggtgc ctgccaggtt cccagtggct ggggtcacca ggtctgaaga gag atg 116
Met
tgc tgg ctg cgg gca tgg ggc cag atc ctc ctg cca gtt ttc ctc tcc 164
Cys Trp Leu Arg Ala Trp Gly Gln Ile Leu Leu Pro Val Phe Leu Ser
-25 -20 -15 -10
ctc ttt ctc atc caa ttg ctt atc agc ttc tca gag aat ggt ttt atc 212
Leu Phe Leu Ile Gln Leu Leu Ile Ser Phe Ser Glu Asn Gly Phe Ile
-5 1 5
cac agc ccc agg aac aat cag aaa cca aga gat ggg aat gaa gag gaa 260
His Ser Pro Arg Asn Asn Gln Lys Pro Arg Asp Gly Asn Glu Glu Glu
10 15 20
tgt gct gta aag aag agt tgt caa ttg tgc aca gaa gat aag aaa tgt 308
Cys Ala Val Lys Lys Ser Cys Gln Leu Cys Thr Glu Asp Lys Lys Cys
25 30 35
gtt tgg tgt agt gaa gaa aaa gca tgc raa aaa tac tgt ttt ccc tat 356
Val Trp Cys Ser Glu Glu Lys Ala Cys Xaa Lys Tyr Cys Phe Pro Tyr
40 45 50 55
ttc ggt tgt cga ttc agt tct ata tat tgg tta aac tgt aaa gtt gac 404
Phe Gly Cys Arg Phe Ser Ser Ile Tyr Trp Leu Asn Cys Lys Val Asp
60 65 70
atg ttt gga atc atg atg ctt cta ctc att gca gta tta att aca gga 452
Met Phe Gly Ile Met Met Leu Leu Leu Ile Ala Val Leu Ile Thr Gly
75 80 85
ttc gtt tgg tac tgc tgc gcc tat cac ttt tac ctg cag gat ata 497
Phe Val Trp Tyr Cys Cys Ala Tyr His Phe Tyr Leu Gln Asp Ile
90 95 100
tgatgaatag ataattgaaa agagatcctc cagaaagagc agaaggaagt ttcttcaatg 557
gcttccttca ggatttttaat catccttaca gcctctttga gaatgattga acttccaaat 617
tccctgaagt taaaatttta aattctatta aacatttttt c 658

<210> 30
<211> 619
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 74..499

<220>
<221> sig_peptide
<222> 74..145
<223> Von Heijne matrix
score 6.56258453552342
seq LSICLSAVATATG/AE

<220>

<221> polyA_site

<222> 556..619

<400> 30

```

agcgactagg acgcgcctgc gcagaggcgg cagcaccacc ggggttgact ccgggggcgc      60
ggcgaggaga gac atg agg ctg agc tgg ttc cgg gtc ctg aca gta ctg      109
          Met Arg Leu Ser Trp Phe Arg Val Leu Thr Val Leu
                    -20                      -15
tcc atc tgc ctg agc gcc gtg gcc acg gcc acg ggg gcc gag ggc aaa      157
Ser Ile Cys Leu Ser Ala Val Ala Thr Ala Thr Gly Ala Glu Gly Lys
          -10                      -5                      1
agg aag ctg cag atc ggg gtc aag aag cgg gtg gac cac tgt ccc atc      205
Arg Lys Leu Gln Ile Gly Val Lys Lys Arg Val Asp His Cys Pro Ile
5          10          15          20
aaa tcg cgc aaa ggg gat gtc ctg cac atg cac tac acg ggt aag ctg      253
Lys Ser Arg Lys Gly Asp Val Leu His Met His Tyr Thr Gly Lys Leu
          25          30          35
gaa gat ggg aca gag ttt gac agc agc ctg ccc cag aac cag ccc ttt      301
Glu Asp Gly Thr Glu Phe Asp Ser Ser Leu Pro Gln Asn Gln Pro Phe
          40          45          50
gtc ttc tcc ctt ggc aca ggc cag gtc atc aag ggc tgg gac cag ggg      349
Val Phe Ser Leu Gly Thr Gly Gln Val Ile Lys Gly Trp Asp Gln Gly
          55          60          65
ctg ctg ggg atg tgt gag ggg gaa aag cgc aas ykg gtg atc cca tcc      397
Leu Leu Gly Met Cys Glu Gly Glu Lys Arg Xaa Xaa Val Ile Pro Ser
          70          75          80
gag cta ggg tat gga gas cgg gga gct ccc cca aag att cca ggc ggt      445
Glu Leu Gly Tyr Gly Xaa Arg Gly Ala Pro Pro Lys Ile Pro Gly Gly
85          90          95          100
gca acc ctg gtg ttc gar gtg gag ctg ctc aaa ata gag cga cga act      493
Ala Thr Leu Val Phe Glu Val Glu Leu Leu Lys Ile Glu Arg Arg Thr
          105          110          115
gag ctg taaccagact ggggaggggc aggggggagag gcccccatca gggaccagac      549
Glu Leu
tggttcaaaaa aaaaaa caaaa aaacaaaaaac aaacaaaaaaa acacttaaaa gccgaaaaaa      609
aaaaaaaaaa

```

<210> 31

<211> 573

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 65..274

<220>

<221> sig_peptide

<222> 65..136

<223> Von Heijne matrix

score 6.54728084640774

ILLTRPACLGSWA/EI

<220>

<221> polyA_site

<222> 556..573

<220>

<221> polyA_signal

<222> 540..545

<400> 31

```

cactgattct ttggccatcg ctactacaat tttagaattg ctggcaattt tttctgcaca      60

```

27

```

ctgg atg gct tct tcc acc agt gtc tca aca gga caa atc ttg ctg aca      109
    Met Ala Ser Ser Thr Ser Val Ser Thr Gly Gln Ile Leu Leu Thr
                -20                      -15                      -10
aga cct gct tgc ttg ggg tcc tgg gct gag atc cgg tca ccg gtg agg      157
Arg Pro Ala Cys Leu Gly Ser Trp Ala Glu Ile Arg Ser Pro Val Arg
                -5                      1                      5
acc atc tcc atc gcc agc gac ttc cca aca gca cgg gtg agt ctc tgg      205
Thr Ile Ser Ile Ala Ser Asp Phe Pro Thr Ala Arg Val Ser Leu Trp
                10                      15                      20
gtg ccg ccc gca cct ggg atg gtt cct att aag atc tcc ggc tgt gca      253
Val Pro Pro Ala Pro Gly Met Val Pro Ile Lys Ile Ser Gly Cys Ala
                25                      30                      35
aac tgg gcc ttc tca ccg gca tagatgatat cacacatcat ggcaagctca      304
Asn Trp Ala Phe Ser Pro Ala
40                      45
cagccccgcg caaaggcata gccattgaca gcagcgatga ctggattctt gacctgggtg      364
agtggtgccc agtgcttcaa gaacttgctg gagtaacagt cctggaaact caggttctgc      424
atttccttga tatcagctct agtgaaaagg gcggtagttg gtggtggaac cagaaacgga      484
cgccggtgct tggagcgggt cttaaattat atttaaaaaa gtaacttttt gtataaataa      544
aagaaaatgg gacaaaaaaa aaaaaaaaaa                        573

```

<210> 32
 <211> 497
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 51..389

<220>
 <221> sig_peptide
 <222> 51..119
 <223> Von Heijne matrix
 score 6.43297166361141
 seq LGALSGAALGFA/SY

<220>
 <221> polyA_site
 <222> 482..497

<220>
 <221> polyA_signal
 <222> 462..467

```

<400> 32
ctcttggttg gcttgggcgg cggttggtccg gtgcgtcctg ttctacagct atg gcc      56
                                Met Ala
ggg cca gct gca gct ttc cgc cgc ttg ggc gcc ttg tcc gga gct gcg      104
Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly Ala Leu Ser Gly Ala Ala
-20                      -15                      -10
gcc tta ggc ttc gct tcc tac ggg gcg cac ggc gcc maw ttc cca gat      152
Ala Leu Gly Phe Ala Ser Tyr Gly Ala His Gly Ala Xaa Phe Pro Asp
-5                      1                      5                      10
gcc tac ggg aag gag ctg ttt gac aag gcc aac aaa cac cac ttc tta      200
Ala Tyr Gly Lys Glu Leu Phe Asp Lys Ala Asn Lys His His Phe Leu
15                      20                      25
cac agc ctg gcc ctg tta ggg gtg ccc cat tgc aga aag cca ctc tgg      248
His Ser Leu Ala Leu Leu Gly Val Pro His Cys Arg Lys Pro Leu Trp
30                      35                      40
gct ggg tta ttg cta gct tcc gga acg acc tta ttc tgc acc agc ttt      296
Ala Gly Leu Leu Leu Ala Ser Gly Thr Thr Leu Phe Cys Thr Ser Phe
45                      50                      55

```

```

tac tac cag gct ctg agt gga gac ccc agc atc cag act ttg gcc cct      344
Tyr Tyr Gln Ala Leu Ser Gly Asp Pro Ser Ile Gln Thr Leu Ala Pro
60                               65                               70                               75
gcg gga ggg acc ctg cta ctc ttg ggc tgg ctt gcc ttg gct ctt      389
Ala Gly Gly Thr Leu Leu Leu Leu Gly Trp Leu Ala Leu Ala Leu
80                               85                               90
tgagctccct tttgcttaat tactggggtt tctgcgcagt tttttttttt aaagagttgg      449
agtaagaaga ggattaaaaa ggaaaggcaa acaaaaaaaaaa aaaaaaaaaa      497

```

<210> 33
 <211> 484
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 14..352

<220>
 <221> sig_peptide
 <222> 14..82
 <223> Von Heijne matrix
 score 6.43297166361141
 seq LGALSGAAALGFA/SY

<220>
 <221> polyA_site
 <222> 469..484

<220>
 <221> polyA_signal
 <222> 443..448

```

<400> 33
ctgttctaca gct atg gcc ggg cca gct gca gct ttc cgc cgc ttg ggc      49
          Met Ala Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly
          -20                               -15
gcc ttg tcc gga gct gcg gcc tta ggc ttc gct tcc tac ggg gcg cac      97
Ala Leu Ser Gly Ala Ala Ala Leu Gly Phe Ala Ser Tyr Gly Ala His
-10                               -5                               1                               5
ggc gcc maa ttc cca gat gcc tac ggg aag gag ctg ttt gac aag gcc      145
Gly Ala Xaa Phe Pro Asp Ala Tyr Gly Lys Glu Leu Phe Asp Lys Ala
10                               15                               20
aac aaa cac cac tty tta cac agc ctg gcc ctg tta ggg gtg ccc cat      193
Asn Lys His His Phe Leu His Ser Leu Ala Leu Leu Gly Val Pro His
25                               30                               35
tgc aga aag cca ctc tgg gct ggg tta ttg cta gct tcc gga acg acc      241
Cys Arg Lys Pro Leu Trp Ala Gly Leu Leu Leu Ala Ser Gly Thr Thr
40                               45                               50
tta tty tgc acc agc ttt tac tac cag gct ytg agt gga gac ccc agc      289
Leu Phe Cys Thr Ser Phe Tyr Tyr Gln Ala Leu Ser Gly Asp Pro Ser
55                               60                               65
atc cag act ttg gcc cct gcg gga ggg acc ctg cta ctc ttg ggc tgg      337
Ile Gln Thr Leu Ala Pro Ala Gly Gly Thr Leu Leu Leu Gly Trp
70                               75                               80                               85
ctt gcc ttg gct ctt tgagctccct tttgattaat tactggggtt tctgggcagt      392
Leu Ala Leu Ala Leu
90
tttttttttt taaagagttg gagtaagaag aggattaaaa aggaaaggca aataaacttt      452
ggagtctttg ttcattcaaaa aaaaaaaaaa aa      484

```

<210> 34
 <211> 739

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 111..272

<220>
<221> sig_peptide
<222> 111..161
<223> Von Heijne matrix
score 6.48133858229514
LTFAQLLFATVLG/IA

<220>
<221> polyA_site
<222> 718..739

<220>
<221> polyA_signal
<222> 688..693

<400> 34
aattgtagac gtcagtgttg gtgttaacat ttcagatttg aagagatttc ctgggtgtgg 60
agtgtgactt tccaaaacca gcttttcctt gagctgtatt tggtgcagca atg ttt 116
Met Phe
agg aga ttg act ttt gca caa ctg ctt ttt gcc act gtc ctt gga att 164
Arg Arg Leu Thr Phe Ala Gln Leu Leu Phe Ala Thr Val Leu Gly Ile
-15 -10 -5 1
gct gga gga gta tat att ttt caa cca gta ttt gaa cag tat gcc aaa 212
Ala Gly Gly Val Tyr Ile Phe Gln Pro Val Phe Glu Gln Tyr Ala Lys
5 10 15
gat cag aag gaa tta aaa gaa aag atg cag ttg gta caa gaa tca gaa 260
Asp Gln Lys Glu Leu Lys Glu Lys Met Gln Leu Val Gln Glu Ser Glu
20 25 30
gag aag aaa agt taatactaca tggagttagg cctggcgagc tggctcacgc 312
Glu Lys Lys Ser
35
ctgtaatccc agcacttttg gaggccgagg cgggtggatc aggtgggtcag gagttcaaga 372
ccagcctgac caacatgggtg aaaccctgtc tctactaaaa atacaaaaat cagccaggct 432
tggtggcatg cgccgtgtaat ccagctact tgggaggctg aggcaggagc atcacttgaa 492
cctgggaggc agaggtggca gtgagtcaag atcacgctgc tgcaactccag cctgggtgac 552
agagcgagac tccatctaaa aaaacaaaaa gcaaaaaaac ccacaaatac ctcatggaga 612
tgaactgtaa taattgctta aagttccatt taattatggt aactctaata tagcaaaaac 672
atagatgtac ttaaaaataa atcatggata atgatttttt aacccaaaaa aaaaaaaaaa 732
tgtgaaa 739

<210> 35
<211> 432
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 47..295

<220>
<221> sig_peptide
<222> 47..91
<223> Von Heijne matrix
score 6.32210467320472
seq LFLSLPVLVVVLS/IV

<220>

<221> polyA_site

<222> 420..432

<220>

<221> polyA_signal

<222> 395..400

<400> 35

```

aaccaagccc tccagcaagg attcagagtg cccctccggc ctgcc atg agg ctc      55
                                   Met Arg Leu
                                   -15
ttc ctg tcg ctt ccg gtc ctg gtg gtg gtt ctg tcg atc gtc ttg gaa      103
Phe Leu Ser Leu Pro Val Leu Val Val Val Leu Ser Ile Val Leu Glu
      -10                        -5                        1
ggc cca gcc cca gcc cag ggg acc cca gac gtc tcc agt gcc ttg gat      151
Gly Pro Ala Pro Ala Gln Gly Thr Pro Asp Val Ser Ser Ala Leu Asp
5      10      15      20
aag ctg aag gag ttt gga aac aca ctg gag gac aag gct cgg gaa ctc      199
Lys Leu Lys Glu Phe Gly Asn Thr Leu Glu Asp Lys Ala Arg Glu Leu
      25      30      35
atc agc cgc atc aaa cag agt gaa ctt tct gcc aag atg cgg gag tgg      247
Ile Ser Arg Ile Lys Gln Ser Glu Leu Ser Ala Lys Met Arg Glu Trp
      40      45      50
ttt tca gag aca ttt cag aaa gtg aag gat aaa ctc aag att gac tca      295
Phe Ser Glu Thr Phe Gln Lys Val Lys Asp Lys Leu Lys Ile Asp Ser
      55      60      65
tgaggacctg aagggtgaca tcccaggagg ggcctctgaa atttcccaca cccagcgcc      355
tgtgctgagg actccctcca tgtggcccca ggtgccacca ataaaaatcc tacagaaaac      415
taawaaaaaa aaaaaaa      432

```

<210> 36

<211> 839

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 66..359

<220>

<221> sig_peptide

<222> 66..194

```

<223> Von Heijne matrix
      score 6.07207487250823
      seq FLIIYVALLRVSEC/LP

```

<220>

<221> polyA_site

<222> 824..839

<220>

<221> polyA_signal

<222> 805..810

<400> 36

```

gggcgcgagg cggccaccgt ggagagcaga ggcgcggcggc tggaagctgc taagtcagag      60
ccgcg atg ttc cgg att gag ggc ctc gcg ccg aag ctg gac ccg gag gag      110
      Met Phe Arg Ile Glu Gly Leu Ala Pro Lys Leu Asp Pro Glu Glu
      -40      -35      -30
atg aaa cgg aag atg cgc gag gat gtg atc tcc tcc ata cgg aac ttt      158
Met Lys Arg Lys Met Arg Glu Asp Val Ile Ser Ser Ile Arg Asn Phe
      -25      -20      -15

```

```

ctc atc tac gtg gcc ctc ctg cga gtc agt gag tgt ctc ccg ggc tgt      206
Leu Ile Tyr Val Ala Leu Leu Arg Val Ser Glu Cys Leu Pro Gly Cys
      -10                      -5                      1
gac tgt gat acc agc ggg gag ctc acc gac ggg cac ccc tta act cta      254
Asp Cys Asp Thr Ser Gly Glu Leu Thr Asp Gly His Pro Leu Thr Leu
5                      10                      15                      20
agg ggt cat cgg ggc ctt cga act gag ctg aac ggt agc ggg gag caa      302
Arg Gly His Arg Gly Leu Arg Thr Glu Leu Asn Gly Ser Gly Glu Gln
      25                      30                      35
gga ggc tcc att tat ctt aaa gaa att gga cag cat atg aag aca gga      350
Gly Gly Ser Ile Tyr Leu Lys Glu Ile Gly Gln His Met Lys Thr Gly
      40                      45                      50
cat cac ata tgaatgcacg atatgaagag cctgggttaca gtttcgactc      399
His His Ile
      55
ctctctgcaa gtgaataggc ccagaaaggt gtaagagact ctttgaatgg acataaaaatt      459
ctgcttggtta agaacaagtt tggctctggt aactgacctt caaagctaaa atataaaaact      519
atctgggaag tatgaaacga tgtctcgtga tctggtgtac ccttatccct gtgacgtttg      579
gcctctgaca atactggtat aattgtaaat aatgtcaaac tccgttttct agcaagtatt      639
aagggagctg tgtctgaaat ggcactgtct tgtcagtcac ttctgtttac ctttttcttc      699
tgcccagagt gtatttgtga aragtctctt atattatgtt ttgtggaaat cagcacacaa      759
ccacaatgac atttaagcac aggatcatta ttagtctatg tttttaataa acatatcaat      819
tatgaaaaaa aaaaaaaaaa      839

```

<210> 37
 <211> 1114
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 117..362

<220>
 <221> sig_peptide
 <222> 117..245
 <223> Von Heijne matrix
 score 5.65876793443964
 VVSFALIATLVYA/LF

<220>
 <221> polyA_site
 <222> 1083..1114

```

<400> 37
ataaggggac gtctagtggg ttgcccggga ggggtggcgg gagcgggcct ggaaataatc      60
tgtcctctgt cgccgggaac tggcgaggta gttccttcgc ggtggagaga cctgga atg      119
                                   Met
gcc aaa tat caa ggt gaa gtt caa agt ttg aaa ctg gat gat gat tca      167
Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp Ser
      -40                      -35                      -30
gtt ata gaa gga gta agc gac caa gta ctt gtg gca gtt gtg gtc agt      215
Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val Ser
      -25                      -20                      -15
ttc gct ttg att gct acc ctg gta tat gca ctt ttc aga aat gta cat      263
Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val His
-10                      -5                      1                      5
caa aac att cac cca gaa aac cag gag cta gta agg gta ctt cga gaa      311
Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg Glu
      10                      15                      20
cag ctt caa aca gaa cag gat gca cct gct gac tcg aca gca gtt cta      359
Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val Leu
      25                      30                      35

```



```

cac tgacatgtac tgtcccatct gcctgcacca agcctccttc ccggtggaga      412
His
ccaactgtgg acatcttttt tgtggtgcct gcattattgc ttaactggcga tatggttcat      472
ggcttggggc aatcagttgt ccaatctgta gacaaacgag acatggccac attgcattgt      532
ccagaacagc ttagaccatg acagtttagca tcgaagccac ctgaggaggg aggcagtaac      592
cttactccta acagtatttg gtgaagatga tcagtctcag gatgttctga gattgcatca      652
ggatattaat gattataacc ggagattctc agggcaaccc agatctgtaa gtaatgctaa      712
agcatgttca aagtttagagg aagacacatt tcttctcttt tgtaaagtga ggtttaccaa      772
caagtattct ttgactatga gaaatcttgg ccaggcacag tagctaacgc ctataatcct      832
agcacttttg gaggccaagg caggtggatc acttgagccc aggagtttga gaccagcctt      892
ggaaacatga tgaaacccca tctctagaaa aaacacaaa aaattggaca agagtgttgg      952
cacatgcctg tagtccttgc ttcttgggag gctgaaatgg gaggatcacc tgagcccagg     1012
aggttgaggc tatagtgagc catgatcgca ctattgcact cccacctggg tggcagtga      1072
acccttcctc aaaaaacaag aaaagaaaaa aaaaaaaaaa aa                      1114

```

```

<210> 38
<211> 1112
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 117..362

```

```

<220>
<221> sig_peptide
<222> 117..245
<223> Von Heijne matrix
      score 5.65876793443964
      seq VVSFALIATLVYA/LF

```

```

<220>
<221> polyA_site
<222> 1097..1112

```

```

<220>
<221> polyA_signal
<222> 1089..1094

```

```

<400> 38
ataaggggac gtctagtggg ttgcccggga ggggtggcgg gagcggtcct ggaaataatc      60
tgtcctctgt cgccgggaac tggcgaggta gttccttcgc ggtggagaga cctgga atg      119
                                     Met
gcc aaa tat caa ggt gaa gtt caa agt ttg aaa ctg gat gat gat tca      167
Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp Ser
      -40                               -35                -30
ggt ata gaa gga gta agc gac caa gta ctt gtg gca gtt gtg gtc agt      215
Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val Ser
      -25                               -20                -15
ttc gct ttg att gct acc ctg gta tat gca ctt ttc aga aat gta cat      263
Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val His
      -10                               -5                   1                   5
caa aac att cac cca gaa aac cag gag cta gta agg gta ctt cga gaa      311
Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg Glu
      10                               15                   20
cag ctt caa aca gaa cag gat gca cct gct gac tcg aca gca gtt cta      359
Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val Leu
      25                               30                   35
cac tgacatgtac tgtcccatct gcctgcacca agcctccttc ccggtggaga      412
His
ccaactgtgg acatcttttt tgtggtgcct gcattattgc ttaactggcga tatggttcat      472
ggcttggggc aatcagttgt ccaatctgta gacaaacgag acatggccac attgcattgt      532
ccagaacagc ttagaccatg acagtttagca tcgaagccac ctgaggaggg aggcagtaac      592

```


gaaacatgat	gaaaccccat	ctctagaaaa	aacaccaaaa	aattggacaa	gagtgttggc	952
acatgcctgt	agtcctgtgt	tcttgggagg	ctgaaatggg	aggatcacct	gagcccagga	1012
ggttgaggct	atagttagcc	atgatcgcac	tattgcactc	ccacctgggt	ggcagtgaga	1072
cccttcctca	aaaaacaaga	aaagaaaaaa	aaaaaaaaaa			1112

<210> 40
 <211> 1112
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 117..362

<220>
 <221> sig_peptide
 <222> 117..245
 <223> Von Heijne matrix
 score 5.65876793443964
 seq VVSFALIATLVYA/LF

<220>
 <221> polyA_site
 <222> 1097..1112

<220>
 <221> polyA_signal
 <222> 1089..1094

<400> 40	
ataaggggac	gtctagtggg
ttgcccggga	gggggtggcgg
gagcggtcct	ggaaataatc
60	
tgctcctctgt	cgccgggaac
tggcgaggta	gttccttcgc
ggtggagaga	cctgga atg
119	Met

gcc aaa tat caa ggt gaa gtt caa agt ttg aaa ctg gat gat gat tca	167
Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp Ser	
-40 -35 -30	

ggt ata gaa gga gta agc gac caa gta ctt gtg gca gtt gtg gtc agt	215
Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val Ser	
-25 -20 -15	

ttc gct ttg att gct acc ctg gta tat gca ctt ttc aga aat gta cat	263
Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val His	
-10 -5 1 5	

caa aac att cac cca gaa aac cag gag cta gta agg gta ctt cga gaa	311
Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg Glu	
10 15 20	

cag ctt caa aca gaa cag gat gca cct gct gac tcg aca gca gtt cta	359
Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val Leu	
25 30 35	

cac tgacatgtac	tgtcccatct	gcctgcacca	agcctccttc	ccggtggaga	412
His					

ccaactgtgg	acatcttttt	tgtggtgcct	gcattattgc	ttactggcga	tatggttcat	472
ggcttggggc	aatcagttgt	ccaatctgta	gacaaacgag	acatggccac	attgcattgt	532
ccagaacagc	ttagaccatg	acagttagca	tcgaagccac	ctgaggaggg	aggcagtaac	592

cttactccta	acagtatttg	gtgaagatga	tcagtctcag	gatgttctga	gattgcatca	652
ggatattaat	gattataacc	ggagattctc	agggcaaccc	agatctgtaa	gtaattgctaa	712
agcatgttca	aagttagagg	aagacacatt	tcttctcttt	tgtaaagtga	ggtttaccac	772

caagtattct	ttgactatga	gaaatcttgg	ccaggcacag	tagtaacgcc	tataatccta	832
gcactttggg	aggccaaggc	aggtggatca	cttgagccca	ggagtttgag	accagccttg	892
gaaacatgat	gaaaccccat	ctctagaaaa	aacaccaaaa	aattggacaa	gagtgttggc	952

acatgcctgt	agtcctgtgt	tcttgggagg	ctgaaatggg	aggatcacct	gagcccagga	1012
ggttgaggct	atagttagcc	atgatcgcac	tattgcactc	ccacctgggt	ggcagtgaga	1072
cccttcctca	aaaaacaaga	aaagaaaaaa	aaaaaaaaaa			1112

<210> 41
 <211> 508
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 17..313

<220>
 <221> sig_peptide
 <222> 17..100
 <223> Von Heijne matrix
 score 5.67441405983879
 seq LGYLVLSEGAFLA/SX

<220>
 <221> polyA_site
 <222> 493..508

<220>
 <221> polyA_signal
 <222> 468..473

<400> 41
 agcaccgaag actgacg atg act tct gca ctg acc cag ggg ctg gag cga atc 52
 Met Thr Ser Ala Leu Thr Gln Gly Leu Glu Arg Ile
 -25 -20
 cca gac cag ctg ggc tac ctg gta ctg agt gaa ggt gca gtg ctg gcg 100
 Pro Asp Gln Leu Gly Tyr Leu Val Leu Ser Glu Gly Ala Val Leu Ala
 -15 -10 -5
 tca tyt ggg gac ctg gag aat gat gag cag gca gcc agt gcc atc tyt 148
 Ser Xaa Gly Asp Leu Glu Asn Asp Glu Gln Ala Ala Ser Ala Ile Xaa
 1 5 10 15
 gag ctg gtc agc aca gcc tgc ggt ttc cgg ctg cac cgc ggc atg aat 196
 Glu Leu Val Ser Thr Ala Cys Gly Phe Arg Leu His Arg Gly Met Asn
 20 25 30
 gtg ccc ttc aag cgc ctg tct gtg gtc ttt gga gaa cac aca ctg ctg 244
 Val Pro Phe Lys Arg Leu Ser Val Val Phe Gly Glu His Thr Leu Leu
 35 40 45
 gtg acg gtg tca gga cag agg gtg ttt gtg gtg aag agg cag aac cga 292
 Val Thr Val Ser Gly Gln Arg Val Phe Val Val Lys Arg Gln Asn Arg
 50 55 60
 ggt cgg gag ccc att gat gtc tgagcctgcc ggagggcgag ggtcggagaa 343
 Gly Arg Glu Pro Ile Asp Val
 65 70
 gcggattggg cttctgtct gccacctcc caccctacc tggacgggcc caggcttggg 403
 gactctgagc tgtgttaagg agaacaagg caaggagacc tccctttgtg ctccctcact 463
 ccctaataaa catgagctctg atgttctcca aaaaaaaaaa aaaaa 508

<210> 42
 <211> 623
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 69..365

<220>
 <221> sig_peptide
 <222> 69..152
 <223> Von Heijne matrix

score 5.54338579743238
seq LGYLVLSEGAFLA/SS

<220>

<221> polyA_site

<222> 608..623

<220>

<221> polyA_signal

<222> 582..587

<400> 42

```
acgtgaccgg ggcctgaagc cggaagctac ctatctggta gggagctccc ccagcaccga      60
agactgcg atg act tct gca ctg acc cag ggg ctg gag cga atc cca gac      110
      Met Thr Ser Ala Leu Thr Gln Gly Leu Glu Arg Ile Pro Asp
                -25                -20                -15
cag ctc ggc tac ctg gta ctg agt gaa ggt gca gtg ctg gcg tca tct      158
Gln Leu Gly Tyr Leu Val Leu Ser Glu Gly Ala Val Leu Ala Ser Ser
                -10                -5                1
ggg gac ctg gag aat gat gag cag gca gcc agt gcc atc tyt gag ctg      206
Gly Asp Leu Glu Asn Asp Glu Gln Ala Ala Ser Ala Ile Xaa Glu Leu
                5                10                15
gtc agc aca gcc tgc ggt ttc cgg ctg cac cgc ggc atg aat gtg ccc      254
Val Ser Thr Ala Cys Gly Phe Arg Leu His Arg Gly Met Asn Val Pro
                20                25                30
ttc aag cgc ctg tct gtg gtc ttt gga gaa cac aca ctg ctg gtg acg      302
Phe Lys Arg Leu Ser Val Val Phe Gly Glu His Thr Leu Leu Val Thr
                35                40                45                50
gtg tca gga cag agg gtg ttt gtg gtg aag agg cag aac cga ggt cgg      350
Val Ser Gly Gln Arg Val Phe Val Val Lys Arg Gln Asn Arg Gly Arg
                55                60                65
gag ccc att gat gtc tgagcctgcc ggagggcgag ggtcggagaa gcggattggg      405
Glu Pro Ile Asp Val
                70
tcctgggcct ctgtgatgag gcaggcacac ctgtcgggtct tggcttgctg ctagaactag      465
ggccttytgs tcgcccacct cccaccccta cctggacggg cccaggcttg gggactytga      525
gctgtgttaa ggagaacaag ggcaaggaga cctccctttg tgctccctca ctccctaata      585
aacatgagtc tgatgtttctc cgaaaaaaaa aaaaaaaaaa      623
```

<210> 43

<211> 1166

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 2..559

<220>

<221> sig_peptide

<222> 2..46

<223> Von Heijne matrix

score 5.34456047740956

seq VMAASMARGGVSA/RV

<220>

<221> polyA_site

<222> 1151..1166

<400> 43

```
g atg cat gtc atg gcc gcc tcc atg gcc cgg gga ggc gtg agt gcc agg      49
Met His Val Met Ala Ser Met Ala Arg Gly Gly Val Ser Ala Arg
      -15                -10                -5                1
```

gtt cta ctg cag gct gcc agg ggc acc tgc tgg aac aga cct ggg ggc	97
Val Leu Leu Gln Ala Ala Arg Gly Thr Xaa Trp Asn Arg Pro Gly Gly	
5 10 15	
act tcc ggg tcg ggg gag ggg gtg gcg ctg ggg aca acc aga aag ttt	145
Thr Ser Gly Ser Gly Glu Gly Val Ala Leu Gly Thr Thr Arg Lys Phe	
20 25 30	
caa gcg aca kgc tcg cgc ccg gct gga gag gag gac gcg ggc ggc ccg	193
Gln Ala Thr Xaa Ser Arg Pro Ala Gly Glu Glu Asp Ala Gly Gly Pro	
35 40 45	
gag cgg ccc ggg gac gtg gtg aac gtg gtg ttc gta gac cgc tca ggc	241
Glu Arg Pro Gly Asp Val Val Asn Val Val Phe Val Asp Arg Ser Gly	
50 55 60 65	
cag cgg atc cca gtg agt ggc aga gtc ggg gac aat gtt ctt cac ctg	289
Gln Arg Ile Pro Val Ser Gly Arg Val Gly Asp Asn Val Leu His Leu	
70 75 80	
gcc cag cgc cac ggg gtg gac ctg gaa ggg gcc tgt gaa gcc tcc ctg	337
Ala Gln Arg His Gly Val Asp Leu Glu Gly Ala Cys Glu Ala Ser Leu	
85 90 95	
gcc tgc tcc acc tgc cat gtg tat gtg agt gaa gac cac ctg gat ctc	385
Ala Cys Ser Thr Cys His Val Tyr Val Ser Glu Asp His Leu Asp Leu	
100 105 110	
ctg cct cct ccc gag gag agg gaa gac gac atg cta gac atg gcc ccc	433
Leu Pro Pro Pro Glu Glu Arg Glu Asp Asp Met Leu Asp Met Ala Pro	
115 120 125	
ctc ctc cag gag aac tcg cgg ctg ggc tgc cag att gtg ctg aca ccg	481
Leu Leu Gln Glu Asn Ser Arg Leu Gly Cys Gln Ile Val Leu Thr Pro	
130 135 140 145	
gag ctg gaa rga gcg gaa ttc acc ctg ccc aag atc acc agg aac ttc	529
Glu Leu Glu Xaa Ala Glu Phe Thr Leu Pro Lys Ile Thr Arg Asn Phe	
150 155 160	
tac gtg gat ggc cat gtc ccc aag ccc cac tgacatgaac acctggacca	579
Tyr Val Asp Gly His Val Pro Lys Pro His	
165 170	
ttccacattg ccatggcccc agggcccaga ttgagggaat agccaggtgc cagccctgcc	639
cagagtgcgg acaggcccgg gagagacgtg gaagcccctg tgaaggacaa caccctgtct	699
tgggagagag tcccatgtcc aggcctctgtt ggggacaggg cccctagtgg ggtggccttc	759
cccaggcccc tgagaatcag gggtttgagta ggagtggact catattggag ctgcaataaa	819
tcgataacac agggcccagc atgtttgagtg tccttggggg acagatctgg ggtctacagg	879
tggctcacac ctgtaatcct agcacttttg gaagccaaga tgggaggatc acttgaggcc	939
aggagcttaa gaccatcctg ggcaacatgg cgagacctcg tctctataaa aacagtaaaa	999
attagcgtgg agtgggtgga agtgcctgta gtcccagcta ctggggagac tgaggtggga	1059
ggatcactta gccaggggga tttaggctgc agtggagcgg tgatcaggcc attccactcc	1119
agcctgggtg acagagttag gccctgtctc taataaaaaa aaaaaaa	1166

<210> 44
 <211> 649
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 18..386

<220>
 <221> sig_peptide
 <222> 18..161
 <223> Von Heijne matrix
 score 5.26064226788865
 seq LLMLLLFLSELQY/YL

<220>
 <221> polyA_site
 <222> 634..649

<220>

<221> polyA_signal

<222> 611..616

<400> 44

```

ctttccggcc ggtcccc atg gag gcg ctg ggg aag ctg aag cag ttc gat      50
                Met Glu Ala Leu Gly Lys Leu Lys Gln Phe Asp
                -45                                -40

gcc tac ccc aag act ttg gag gac ttc cgg gtc aag acc tgc ggg ggc      98
Ala Tyr Pro Lys Thr Leu Glu Asp Phe Arg Val Lys Thr Cys Gly Gly
                -35                                -30                                -25

gcc acc gtg acc att gtc agt ggc ctt ctc atg ctg cta ctg ttc ctg      146
Ala Thr Val Thr Ile Val Ser Gly Leu Leu Met Leu Leu Leu Phe Leu
                -20                                -15                                -10

tcc gag ctg cag tat tac ctc acc acg gag gtg cat cct gag ctc tac      194
Ser Glu Leu Gln Tyr Tyr Leu Thr Thr Glu Val His Pro Glu Leu Tyr
-5                                1                                5                                10

gtg gac aag tcg cgg gga gat aaa ctg aag atc aac atc gat gta ctt      242
Val Asp Lys Ser Arg Gly Asp Lys Leu Lys Ile Asn Ile Asp Val Leu
                15                                20                                25

ttt ccg cac atg cct tgt gcc tat ctg agt att gat gcc atg gat gtg      290
Phe Pro His Met Pro Cys Ala Tyr Leu Ser Ile Asp Ala Met Asp Val
                30                                35                                40

gcc gga gaa cag car ctg gat gtg gaa cac aac ctg ttc aag caa cga      338
Ala Gly Glu Gln Gln Leu Asp Val Glu His Asn Leu Phe Lys Gln Arg
                45                                50                                55

cta gat aaa gat ggc atc ccc gtg agc tca gag gct gag cgg cat gmt      386
Leu Asp Lys Asp Gly Ile Pro Val Ser Ser Glu Ala Glu Arg His Xaa
60                                65                                70                                75

taatgcagga caccagtga accaacataa tttgctatag gactggatgt ggagtggagg      446
agagcaaagt catggttggt gattccaacg ctttcagcct gagcaaata ga atgacttgag      506
atggagatga ctggaggaga aggttttttt ggtgggaggg aggagtggct attggatatc      566
cmctgctgtg taacaaagta ctccaaaagt tagtagtttg atacaataaa aatgtgtttt      626
ttcagtcaaa aaaaaaaaaa aaa                                           649

```

<210> 45

<211> 836

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 87..317

<220>

<221> sig_peptide

<222> 87..173

```

<223> Von Heijne matrix
      score 5.19818357425197
      seq LLLKLQRLPQAEP/VE

```

<220>

<221> polyA_site

<222> 821..836

<220>

<221> polyA_signal

<222> 805..810

<400> 45

```

gagtgtgggg ccagccgtgg aggtccagg tgttctctct gccccagcag agcccggcag      60
gagccccaac aggaagccag cgcggc atg gct gcc acc gac ttc gtg cag gag      113

```

Met Ala Ala Thr Asp Phe Val Gln Glu
-25

atg cgc gcc gtg ggc gag agg ctg ctg ctc aag ctg cag aga ctg ccc 161
Met Arg Ala Val Gly Glu Arg Leu Leu Leu Lys Leu Gln Arg Leu Pro
-20 -15 -10 -5

cag gct gag ccc gtg gag atc gtg gcc ttc tca gtc atc atc ctt ttc 209
Gln Ala Glu Pro Val Glu Ile Val Ala Phe Ser Val Ile Ile Leu Phe
1 5 10

aca gct act gtt ctg ctg ttg ctg ctg ata gcc tgc agc tgc tgc tgc 257
Thr Ala Thr Val Leu Leu Leu Leu Leu Ile Ala Cys Ser Cys Cys Cys
15 20 25

act cac tgc tgc tgc cct gag cgg aga ggc agg aag gtc cag gtg cag 305
Thr His Cys Cys Cys Pro Glu Arg Arg Gly Arg Lys Val Gln Val Gln
30 35 40

ccg aca cca cca tgacggacgg gcgatggctg aggagaagct ggagaggaga 357
Pro Thr Pro Pro
45

tgccaatgc catgacacag gccatcagcc tggccctgca gcccttacc cccaagacca 417
ggctcccctg gcccagctc tggcccagcc caggtacctg gacactgaca acttgagccc 477
taccaaggaa acaagggtg gtataggtgc aaacctctca tctgccagtg gacactgggt 537
gctggggagt cagctgtttc aaagactggg tcaactgcct gggcttcttc gcctacctgc 597
actttttaac aaaacaagga agtaggggtc cccatacctt gatggagaac agtccccacc 657
tgtgggcaat tggcccttg ggctctgctg atacatgcc aagaggagca aggcaatcag 717
aggggctttg tgcaatagct tctgcatccg agctcccgcc agagcgtgag catgtcagta 777
ttctagtcca gtatttgcca gtttccaagt aaaagctttt gtgaaaaaaaa aaaaaaaaaa 836

<210> 46
<211> 482
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 17..337

<220>
<221> sig_peptide
<222> 17..97
<223> Von Heijne matrix
score 4.95181310301952
seq VLSTLSLVQFSPS/GR

<220>
<221> polyA_site
<222> 467..482

<220>
<221> polyA_signal
<222> 442..447

<400> 46

acttttagctg ctgcag atg ctg ttt gaa gaa gct ctg ccc ctc agc tgc tct 52
Met Leu Phe Glu Glu Ala Leu Pro Leu Ser Cys Ser
-25 -20

gat cct gtg ctt agc act ctt agc ctg gtg cag ttc agc ccc agt gga 100
Asp Pro Val Leu Ser Thr Leu Ser Leu Val Gln Phe Ser Pro Ser Gly
-15 -10 -5 1

agg acc cag gac ctg ctc tct cca ggg gtg gag aac ctg tcg gtg ctg 148
Arg Thr Gln Asp Leu Leu Ser Pro Gly Val Glu Asn Leu Ser Val Leu
5 10 15

gac gtg tcc cct ctg ggc ttg gcc tgc tgt ctg ctc act ctc acc atg 196
Asp Val Ser Pro Leu Gly Leu Ala Cys Cys Leu Leu Thr Leu Thr Met
20 25 30

40

```

tcc tgc cca ggg cct gac cct cct gag ggg ccc ggg acc cag cgt gtg      244
Ser Cys Pro Gly Pro Asp Pro Pro Glu Gly Pro Gly Thr Gln Arg Val
      35                      40                      45
tgg caa ggg gct cta cgg atc cta cag ctc cca gga gcc cca gat ggg      292
Trp Gln Gly Ala Leu Arg Ile Leu Gln Leu Pro Gly Ala Pro Asp Gly
      50                      55                      60                      65
gtt tca cca tat cag cca gtc tgg tct cga act cct gac ctc aag      337
Val Ser Pro Tyr      70 Pro Val Trp Ser Arg Thr Pro Asp Leu Lys
                                75                                80
tgattcgcca gcctcgctcct cccaaagtgc tgggattaca gatgtgagcc accatgccca      397
tggagaatgt attacttttg taatttaaaa aataatagta ataaaaataaa gatctttaca      457
atgttttcaga aaaaaaaaaa aaaaaa      482

```

<210> 47
 <211> 806
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 136..471

<220>
 <221> sig_peptide
 <222> 136..252
 <223> Von Heijne matrix
 score 4.89253405263876
 seq IILFSAIVGFIYG/YV

<220>
 <221> polyA_site
 <222> 792..806

<220>
 <221> polyA_signal
 <222> 775..780

```

<400> 47
ttcgtctctcc cgggcttaga aggcccggt actgacgcgc agtgccagac cttaccctc      60
acggctcctta agtctcggtc gccctcgct cgcagcctgc caccgcgct cagctgccg      120
cctcctcagc cagcc atg ctg gag cat ctg agc tcg ctg ccc acg cag atg      171
      Met Leu Glu His Leu Ser Ser Leu Pro Thr Gln Met
                                -35                                -30
gat tac aag ggc cag aag cta gct gaa cag atg ttt cag gga att att      219
Asp Tyr Lys Gly Gln Lys Leu Ala Glu Gln Met Phe Gln Gly Ile Ile
      -25                      -20                      -15
ctt ttt tct gca ata gtt gga ttt atc tac ggg tac gtg gct gaa cag      267
Leu Phe Ser Ala Ile Val Gly Phe Ile Tyr Gly Tyr Val Ala Glu Gln
      -10                      -5                      1                      5
ttc ggg tgg act gtc tat ata gtt atg gcc gga ttt gct ttt tca tgt      315
Phe Gly Trp Thr Val Tyr Ile Val Met Ala Gly Phe Ala Phe Ser Cys
                                10                      15                      20
ttg ctg aca ctt cct cca tgg ccc atc tat cgc cgg cat cct ctc aag      363
Leu Leu Thr Leu Pro Pro Trp Pro Ile Tyr Arg Arg His Pro Leu Lys
      25                      30                      35
tgg tta cct gtt caa gca cag acg aca aga aac cag ggg aaa gaa aaa      411
Trp Leu Pro Val Gln Ala Gln Thr Thr Arg Asn Gln Gly Lys Glu Lys
      40                      45                      50
tta aga ggc atg cta aaa ata att gag gtt ttc atg att cag cac ctg      459
Leu Arg Gly Met Leu Lys Ile Ile Glu Val Phe Met Ile Gln His Leu
      55                      60                      65
ctt ttg ttt ctg tgagatgagc taaattgctt tcatacccca gataagagct      511
Leu Leu Phe Leu

```

70
 aaaaccacct aatgctctta tggcacagct gtgtatagat ttagttctct ttataacttca 571
 tttctagccc agttgggttt tgatttataat aagtagttta gaccttctct tcataaatctt 631
 gctctgagat ggggaacaga acacacaagt atgaagtttc tttcaggtgt aaataatgaa 691
 aaataaatgc ctcataaatg atagtacaat gtaactatca aagttttata attcattatg 751
 agttaaccat tttaatgttt ccaattaaac ctcatagtgc aaaaaaaaaa aaaaa 806

<210> 48
 <211> 582
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 148..360

<220>
 <221> sig_peptide
 <222> 148..216
 <223> Von Heijne matrix
 score 5.918755252201
 SALMLLPCRPLT/SV

<220>
 <221> polyA_site
 <222> 566..582

<220>
 <221> polyA_signal
 <222> 544..549

<400> 48
 aacgtcatct aggagcacgc agcagcttgg ctaaaagtaa ggggtgctgtg ctgatggccc 60
 tgtgcgcact gacccgckct sygtctcttg aacctggcgc ccccgaccgt cgccgcccct 120
 gccccgagtc tgttccccgc cgcccag atg atg aac aat ggc ctc ctc caa cag 174
 Met Met Asn Asn Gly Leu Leu Gln Gln
 -20 -15
 ccc tct gcc ttg atg ttg ctc ccc tgc cgc cca gtt ctt act tct gtg 222
 Pro Ser Ala Leu Met Leu Leu Pro Cys Arg Pro Val Leu Thr Ser Val
 -10 -5 1
 gcc ctt aat gcc aac ttt gtg tcc tgg aag agt cgt acc aag tac acc 270
 Ala Leu Asn Ala Asn Phe Val Ser Trp Lys Ser Arg Thr Lys Tyr Thr
 5 10 15
 att aca cca gtg aag atg agg aag tct ggg ggc cga gac cac aca ggt 318
 Ile Thr Pro Val Lys Met Arg Lys Ser Gly Gly Arg Asp His Thr Gly
 20 25 30
 gct gga aac gtg cgt agc aac agt agg ccg agt atc caa cgt 360
 Ala Gly Asn Val Arg Ser Asn Ser Arg Pro Ser Ile Gln Arg
 35 40 45
 tgatcataac aaacgggtca ttggcaaggc aggtcgcaac cgctggctgg gcaagaggcc 420
 taacagtggg cgggtggcacc gcaagggggg ctgggctggc cgaaagattc ggccactacc 480
 ccccatgaag agttacgtga agctgccttc tgcttctgcc caaagctgat atccctgtac 540
 tctaataaaa tgcccccccc ccctcaaaaa aaaaaaaaaa aa 582

<210> 49
 <211> 583
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 149..361

<220>

<221> sig_peptide

<222> 149..217

<223> Von Heijne matrix

score 5.918755252201

SALMLLPCRPLT/SV

<220>

<221> polyA_site

<222> 567..583

<220>

<221> polyA_signal

<222> 545..550

<400> 49

```

aacgtcatct aggagcaccg agcagcttgg ctaaaagtaa ggggtgtcgtg ctgatggccc      60
tgtgcgcact gaccgcgcgt ctgcstctct gaacctggcg ccccgaccg tcgccgcccc      120
tgccccgagt ctgttccccg ccgcccag atg atg aac aat ggc ctc ctc caa      172
                               Met Met Asn Asn Gly Leu Leu Gln
                               -20

```

```

cag ccc tct gcc ttg atg ttg ctc ccc tgc cgc cca gtt ctt act tct      220
Gln Pro Ser Ala Leu Met Leu Leu Pro Cys Arg Pro Val Leu Thr Ser
-15                               -10                               -5                               1

```

```

gtg gcc ctt aat gcc aac ttt gtg tcc tgg aag agt cgt acc aag tac      268
Val Ala Leu Asn Ala Asn Phe Val Ser Trp Lys Ser Arg Thr Lys Tyr
      5                               10                               15

```

```

acc att aca cca gtg aag atg agg aag tct ggg ggc cga gac cac aca      316
Thr Ile Thr Pro Val Lys Met Arg Lys Ser Gly Gly Arg Asp His Thr
      20                               25                               30

```

```

ggg gct gga aac gtg cgt agc aac agt agg ccg agt atc caa cgt      361
Gly Ala Gly Asn Val Arg Ser Asn Ser Arg Pro Ser Ile Gln Arg
      35                               40                               45

```

```

tgatcataac aaacgggtca ttggcaaggc aggtcgcaac cgctggctgg gcaagaggcc      421
taacagtggg cgggtggcacc gcaagggggg ctgggctggc cgaaagattc ggccactacc      481
ccccatgaag agttacgtga agctgccttc tgcttctgcc caaagctgat atccctgtac      541
tctaataaaa tgcccccccc cctcaaaaaa aaaaaaaaaa aa      583

```

<210> 50

<211> 746

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 83..358

<220>

<221> sig_peptide

<222> 83..127

<223> Von Heijne matrix

score 4.68058603039206

VLASLLGPTSRS/AA

<400> 50

```

ctttttttcc tttgttccag ggctggagcg gctctgggct ccggaatcgc ccgcagccgg      60
tactgcggga cccactgcgg at atg gct gtc ttg gct gga tcc ctg ttg ggc      112
                               Met Ala Val Leu Ala Gly Ser Leu Leu Gly
                               -15                               -10

```

```

ccc acg agt agg tcg gca gcg ttg ctg ggt ggc agg tgg ctc cag ccc      160
Pro Thr Ser Arg Ser Ala Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro
-5                               1                               5                               10

```

```

cgg gcc tgg ctg ggg ttc cca gac gcc tgg ggc ctc ccc acc ccg cag      208

```

```

Arg Ala Trp Leu Gly Phe Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln
      15                20                25
cag gcc cgg ggc aag gct cgc ggg aat gag tat cag ccg agc aac atc      256
Gln Ala Arg Gly Lys Ala Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile
      30                35                40
aaa cgc aag aac aag cac ggc tgg gtc cgg cgc ctg agc acg cca gcc      304
Lys Arg Lys Asn Lys His Gly Trp Val Arg Arg Leu Ser Thr Pro Ala
      45                50                55
ggc gts cag gtc atc ctt cgc cga atg ctc aag ggc cgc aag tcg ctg      352
Gly Val Gln Val Ile Leu Arg Arg Met Leu Lys Gly Arg Lys Ser Leu
      60                65                70                75
agc cat tgaggatcgc gacgcagtcg gcgggaccct catggaagca tcgccctcgc      408
Ser His
ctcggacctt gcctggcgct atttttgcag ggagctgggg agcaggaacg cctcggacct      468
gagtgtctct catattgttg ggttgaagtc tggatgggag cttgccaaagt cccttttttag      528
gctttttaat taggaagcat ttcgaacctg cgcaacagac caaagaacag tacaaagaac      588
atccgtgtac ccagtaccct gactaccgac tacctacaac ccgtccctgc cccatcctga      648
gttcttttga agctgatctc aggcacgcga ttatttcttc tgtaaataatt tcagaatgta      708
tctctccaag atgagagctc attaaaagac aattacaa      746

<210> 51
<211> 731
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 73..348

<220>
<221> sig_peptide
<222> 73..117
<223> Von Heijne matrix
      score 4.68058603039206
      seq VLAGSLLGPTSRS/AA

<400> 51
gttgttccag ggctggagcg gctctgggct ccggaatcgc ccgcagccgg tactgcggga      60
cccactgcag at atg gct gtc ttg gct gga tcc ctg ttg ggc ccc acg agt      111
      Met Ala Val Leu Ala Gly Ser Leu Leu Gly Pro Thr Ser
      -15                -10                -5
agg tcg gca gcg ttg ctg ggt ggc agg tgg ctc car ccc cgg gcc tgg      159
Arg Ser Ala Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro Arg Ala Trp
      1                5                10
ctg ggg ttc cca gac gcc tgg ggc ctc ccc acc ccg cag cag gcc cgg      207
Leu Gly Phe Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln Ala Arg
      15                20                25                30
ggc aag gct cgc ggg aat gag tat cag ccg agc aat atc aaa cgc aag      255
Gly Lys Ala Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys Arg Lys
      35                40                45
aac aag cac ggc tgg gtc cgc cgc ctg arc acg ccg gcc ggc gtg cag      303
Asn Lys His Gly Trp Val Arg Arg Leu Xaa Thr Pro Ala Gly Val Gln
      50                55                60
gtc atc ctt cgc cga atg ctc aag ggc cgc aag tcg ctg agc cat      348
Val Ile Leu Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser His
      65                70                75
tgaggatcgc gacgcagtcg gcgggaccct catggaagca tcgccctcgc ctcggacctt      408
gcctggcgct atttttgcag ggagctgggg agcaggaacg cctcggacct gagtgtctct      468
catattgttg ggttgaagtc tggatgggag cttgccaaagt cccttttttag gctttttaat      528
taggaagcat ttcgaacctg cgcaacagac caaagaacag tacaaagaac atccgtgtac      588
ccagtaccct gactaccgac tacctacaac ccgtccctgc cccatcctga gttcttttga      648
agctgatctc aggcacgcga ttatttcttc tgtaaataatt tcagaatgta tctctccaag      708
atgagagctc attaaaagat aat      731

```

```
<210> 52
<211> 773
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 22..297
```

```
<220>
<221> sig_peptide
<222> 22..66
<223> Von Heijne matrix
score 4.68058603039206
seq VLAGSLLGPTSRS/AA
```

```
<220>
<221> polyA_site
<222> 758..773
```

```
<220>
<221> polyA_signal
<222> 740..745
```

<400>	52
actgctgggac ccactgcgga t atg gct gtc ttg gct gga tcc ctg ttg ggc	51
	Met Ala Val Leu Ala Gly Ser Leu Leu Gly
-15	-10
ccc acg agt agg tcg gca gcg ttg ctg ggt ggc agg tgg ctc cag ccc	99
Pro Thr Ser Arg Ser Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro	
-5	1 5 10
cgg gcc tgg ctg ggg ttc cca gac gcc tgg ggc ctc ccc acc ccg cag	147
Arg Ala Trp Leu Gly Phe Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln	
15	20 25
cag gcc cgg ggc aag gct cgc ggg aat gar tat cag ccg agc aat atc	195
Gln Ala Arg Gly Lys Ala Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile	
30	35 40
aaa cgc aag aac aag cac gcc tgg gtc cgg cgc ctg arc acg ccg gcc	243
Lys Arg Lys Asn Lys His Gly Trp Val Arg Arg Leu Xaa Thr Pro Ala	
45	50 55
ggc gtg mrg gtc atc ctt cgc cga atg ctc aag ggc cgc aag tcg ctg	291
Gly Val Xaa Val Ile Leu Arg Arg Met Leu Lys Gly Arg Lys Ser Leu	
60	65 70 75
agc cat tgaggatcg cgcgcagtgc gcgggaccct catggaagca tcgccctcgc	347
Ser His	
ctcggacatt gcctggcgct atttttgtag ggagctgggg agcaggaacg cctcggacct	407
gagtgcctct catatttgtg gtttgaagtc tggatgggag cttgccaaagt ccccttttag	467
gctttttaaat taggaagcat ttcgaacctg cgcaacagac caaagaacag tacaaagaac	527
atccgtgtac ccagtacct gactaccgac tacctacaac ccgtccctgc cccatcctga	587
gttctttttga agctgatctc aggcatcgga ttattttcttc tgtaaatat tccagaatgta	647
tctctccaag atgagagctc attaaaaagat aattacaaag cttatcacat ccaaaaagaat	707
tatcaataat tttgaaatat tattaaacgt gtaataaatg ttcaaagttc aaaaaaaaaa	767
aaaaaa	773

```
<210> 53
<211> 740
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 80..355
```

```

<220>
<221> sig_peptide
<222> 80..124
<223> Von Heijne matrix
      score 6.4806443024348
      seq VLAGSLLXPTSR/AA

<400> 53
ataatccttt gttccagggc tggagcggct ctgggctccg gaatcgcccg cagccggtac      60
tgcgggaccc actgcggat atg gct gtc ttg gct gga tcc ctg ttg grc ccc      112
              Met Ala Val Leu Ala Gly Ser Leu Leu Xaa Pro
              -15                      -10                      -5

acg agt agg tcg gca gcg ttg ctg ggt rgc agg tgg ctc cag ccc cgg      160
Thr Ser Arg Ser Ala Ala Leu Leu Gly Xaa Arg Trp Leu Gln Pro Arg
              1                      5                      10

gcc tgg ctg ggg ttc cca gac gcc tgg ggc ctc ccc acc ccg cag cag      208
Ala Trp Leu Gly Phe Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln
              15                      20                      25

gcc cgg ggc aag gct cgc ggg aat gag tat cag ccg agc aac atc aaa      256
Ala Arg Gly Lys Ala Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys
              30                      35                      40

cgc aag aac aag cac ggc tgg gtc cgg cgc ctg arc acg ccg gcc ggc      304
Arg Lys Asn Lys His Gly Trp Val Arg Arg Leu Xaa Thr Pro Ala Gly
              45                      50                      55                      60

kwr smg ktc atc ctt cgc cga atg ctc aag ggc cgc aag tcg ctg agc      352
Xaa Xaa Xaa Ile Leu Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser
              65                      70                      75

cat tgaggatcgc gacgcagtsq scggggaacc ctcatggaag catcgccctc      405
His
gcctcggacc ttgcctggcg ctatcttttgc agggagctgg ggagcaggaa cgctcggac      465
ctgagtgtctc tccatattgt ggggttgaag tctggatggg agcttgccaa gtcccttttt      525
aggcttttta attaggaagc atttcgaacc tgcgcaacag accaaagaac agtacaaaga      585
acatccgtgt acccagtacc ctgactaccg actacctaca acccgccctt gccccatcct      645
gagttctttt gaagctgacg tcaggcatcg gattatttct tctgtaaata tttcagaatg      705
tatctctcca agatgagagc tcattaaaag acaat      740

<210> 54
<211> 638
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 137..397

<220>
<221> sig_peptide
<222> 137..253
<223> Von Heijne matrix
      score 4.61343345943784
      seq LQGARVLLTGANA/GV

<220>
<221> polyA_site
<222> 623..638

<220>
<221> polyA_signal
<222> 616..621

<400> 54
ggggtgcggt cggggacccg gcaggaggcg gccgagaaga gaggaccgtg ggggcgttcg      60

```

```

cgtggctccc agcccgggac cccacccccg ctggacagtg ggggaaactg aggcctgagc 120
gggcccacac aggacc atg aag gtg ctt ctc ctc aca ggg ctg ggg gcc ctg 172
          Met Lys Val Leu Leu Leu Thr Gly Leu Gly Ala Leu
          -35                      -30

ttc ttc gcc tat tat tgg gat gac aac ttc gac cca gcc agc ctc cag 220
Phe Phe Ala Tyr Tyr Trp Asp Asp Asn Phe Asp Pro Ala Ser Leu Gln
          -25                      -20                      -15

gga gcg cga gtg ctg ctg aca ggg gcc aac gct ggt gtt ggt gag gag 268
Gly Ala Arg Val Leu Leu Thr Gly Ala Asn Ala Gly Val Gly Glu Glu
          -10                      -5                      1                      5

ctg gcc tat cac tac gcg cgt ctg ggc tcc cac ctg gtg ctc act gcc 316
Leu Ala Tyr His Tyr Ala Arg Leu Gly Ser His Leu Val Leu Thr Ala
          10                      15                      20

cac act gag gct ctc ctg cag aag gca cgg tgg ctc acg ctt gta gtc 364
His Thr Glu Ala Leu Leu Gln Lys Ala Arg Trp Leu Thr Leu Val Val
          25                      30                      35

agc act ttg gga ggc cga gga aag tgg atc acc tgagggtcagg agttcaagac 417
Ser Thr Leu Gly Gly Arg Gly Lys Trp Ile Thr
          40                      45

cagcctggcc aacatggcga aacctggtct ctactaagaa gacaaaaatt aggccaggca 477
tggtggctca tgctgtaat cctagcactt tgggaggcca aggcggatgg atcacttgag 537
gtcaggagtt caaaaccagc ctggccaaca tgggtgaaacc ctgtctctac taaaaatata 597
aaaaataaaa ataaaataaa taaacaaaaa aaaaaaaaaa a 638

<210> 55
<211> 577
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 18..392

<220>
<221> sig_peptide
<222> 18..62
<223> Von Heijne matrix
      score 4.60228634566274
      seq LLTHNLLSSHVRG/VG

<220>
<221> polyA_site
<222> 562..577

<220>
<221> polyA_signal
<222> 545..550

<400> 55
tttttttgtg cggcgac atg aaa ctg ctt acc cac aat ctg ctg agc tcg 50
          Met Lys Leu Leu Thr His Asn Leu Leu Ser Ser
          -15                      -10                      -5

cat gtg cgg ggg gtg ggg tcc cgt ggc ttc ccc ctg cgc ctc cag gcc 98
His Val Arg Gly Val Gly Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala
          1                      5                      10

acc gag gtc cgt atc tgc cct gtg gaa ttc aac ccc aac ttc gtg gcg 146
Thr Glu Val Arg Ile Cys Pro Val Glu Phe Asn Pro Asn Phe Val Ala
          15                      20                      25

cgt atr ata cct aaa gtg gag tgg tcg gcg ttc ctg gag gcg gcc gat 194
Arg Xaa Ile Pro Lys Val Glu Trp Ser Ala Phe Leu Glu Ala Ala Asp
          30                      35                      40

aac ttg cgt ctg atc cag gtg ccg aaa ggg ccg gtt gag gga tat gag 242
Asn Leu Arg Leu Ile Gln Val Pro Lys Gly Pro Val Glu Gly Tyr Glu

```

47

45	50	55	60	
gag aat gag gag ttt ctg agg acc atg cac cac ctg ctg ctg gag gtg				290
Glu Asn Glu Glu Phe Leu Arg Thr Met His His Leu Leu Leu Glu Val				
	65	70	75	
gaa gtg ata gag ggc acc ctg cag tgc ccg gaa tct gga cgt atg ttc				338
Glu Val Ile Glu Gly Thr Leu Gln Cys Pro Glu Ser Gly Arg Met Phe				
	80	85	90	
ccc atc agc cgc ggg atc ccm amc atg ctg ctg agt gaa gag gaa act				386
Pro Ile Ser Arg Gly Ile Pro Xaa Met Leu Leu Ser Glu Glu Glu Thr				
	95	100	105	
gag agt tgattgtgsc aggcgccagt ttttcttggt atgactgttt atttttgttg				442
Glu Ser				
110				
atctataccc tgtttccgaa ttctgccgtg tgtatcccca acccttgacc caatgacacc				502
aaacacagtg tttttgagct cgggtattata tttttttttc tcattaaagg tttaaaacca				562
aaaaaaaaaa aaaaa				577

<210> 56
 <211> 521
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 82..360

<220>
 <221> sig_peptide
 <222> 82..147
 <223> Von Heijne matrix
 score 4.53239781724629
 seq FFLITIALGTKT/ES

<220>
 <221> polyA_site
 <222> 506..521

<220>
 <221> polyA_signal
 <222> 486..491

<400> 56	
attttctcaa caattcctca ccgcaggagc ctctgaagct cccaccaggc cagctctcct	60
cccacaacag cttcccacag c atg aag atc tcc gtg gct gcc att ccc ttc	111
	Met Lys Ile Ser Val Ala Ala Ile Pro Phe
	-20 -15
ttc ctc ctc atc acc atc gcc cta ggg acc aag act gaa tcc tcc tca	159
Phe Leu Leu Ile Thr Ile Ala Leu Gly Thr Lys Thr Glu Ser Ser Ser	
	-10 -5 1
cgg gga cct tac cac ccc tca gag tgc tgc ttc acc tac act acc tac	207
Arg Gly Pro Tyr His Pro Ser Glu Cys Cys Phe Thr Tyr Thr Thr Tyr	
5 10 15 20	
aag atc ccg cgt cag cgg att atg gat tac tat gag acc aac agc cag	255
Lys Ile Pro Arg Gln Arg Ile Met Asp Tyr Tyr Glu Thr Asn Ser Gln	
	25 30 35
tgc tcc aag ccc gga att gtc ttc atc acc aaa agg ggc cat tcc gty	303
Cys Ser Lys Pro Gly Ile Val Phe Ile Thr Lys Arg Gly His Ser Val	
	40 45 50
tgt acc aac ccc agt gac aag tgg gtc cag gac tat atc aag gac atg	351
Cys Thr Asn Pro Ser Asp Lys Trp Val Gln Asp Tyr Ile Lys Asp Met	
	55 60 65
aag gag aac tgagtgaccc agaaggggtg gcgaaggcac agctcagaga	400
Lys Glu Asn	

70
cataaagaga agatgccaaag gccccctcct ccacccaccg ctaactctca gccccagtca 460
ccctcttgga gcttcctgcg tttgaattaa agaccactca tgcccaaaaa aaaaaaaaaa 520
a 521

<210> 57
<211> 588
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 93..308

<220>
<221> sig_peptide
<222> 93..218
<223> Von Heijne matrix
score 4.46006149913644
seq LRLEVAAAPGTPA/QP

<220>
<221> polyA_site
<222> 573..588

<220>
<221> polyA_signal
<222> 558..563

<400> 57
aaaagaaatt cccggggcccg gccagcctgt gccgctgtta tgaggagaag caaagcccct 60
tgcagaagca cttgcggggtc tgcagcctac gc atg aat agg ttg gca ggt gtg 113
Met Asn Arg Leu Ala Gly Val
-40
ggc tgg cgg gtg gac tac acc ctg agc tcc agc ctg ctg caa tcc gtg 161
Gly Trp Arg Val Asp Tyr Thr Leu Ser Ser Leu Leu Gln Ser Val
-35 -30 -25 -20
gaa gag ccc atg gtg cac ctg cgg ctg gag gtg gca gct gcc cca ggg 209
Glu Glu Pro Met Val His Leu Arg Leu Glu Val Ala Ala Ala Pro Gly
-15 -10 -5
acc cca gcc cag cct gtt gcc atg tcc ctg tca gca gac aag ttc cag 257
Thr Pro Ala Gln Pro Val Ala Met Ser Leu Ser Ala Asp Lys Phe Gln
1 5 10
gtc ctg ctg gca gaa ctg aag cag gcc cag acc ctg atg agc tcc ctg 305
Val Leu Leu Ala Glu Leu Lys Gln Ala Gln Thr Leu Met Ser Ser Leu
15 20 25
ggc tgaggagaag ggtgttccag gcctgtgtgg agccgcctg ccggtatgga 358
Gly
30
gtcacgccct ctgaactgct cttcgggagg cagccctggt tctaggatgc tgaggccctg 418
gcccggaactc tggcctccca gatccccagc tgcctcaactt ctctcttgag aacttggctc 478
agggtcctctg aggacctttc ccagcattac cttcccttcc cttgaaaggc aattgttggc 538
tgttttcata agcaggaaaa ataaacagaa gtacaaaaaa aaaaaaaaaa 588

<210> 58
<211> 521
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 113..364

<220>

<221> sig_peptide

<222> 113..172

<223> Von Heijne matrix

score 4.37180298395146

SLLLSLPPHQGLT/FS

<220>

<221> polyA_site

<222> 500..521

<400> 58

ttttttacat ggtgttccca cagctgggag gacacccaca tggtcggcgt gcaggatatt 60

tcgctggacc ctagaaaagc caccacgacc tgtgggccat gatgctaccc ca atg gct 118

Met Ala

-20

gct gct gct gtt cct tct ctt ctt ctt tct ctt cct cct cac cag ggg 166

Ala Ala Ala Val Pro Ser Leu Leu Ser Leu Pro Pro His Gln Gly

-15

-10

-5

ctc act ttc tcc aac aaa ata caa cct ttt gga gct caa gga gtc ttg 214

Leu Thr Phe Ser Asn Lys Ile Gln Pro Phe Gly Ala Gln Gly Val Leu

1

5

10

cat ccg gaa cca gga ctg cga gac tgg ctg ctg cca acg tgc tcc aga 262

His Pro Glu Pro Gly Leu Arg Asp Trp Leu Leu Pro Thr Cys Ser Arg

15

20

25

30

caa ttg cga gtc gca ctg ccg gag aag ggg tcc gag ggc agt ctg tgt 310

Gln Leu Arg Val Ala Leu Pro Glu Lys Gly Ser Glu Gly Ser Leu Cys

35

40

45

caa acg cag ctg cca gct act cca tgc ttc ctg cct tcg aat acg gtg 358

Gln Thr Gln Leu Pro Ala Thr Pro Cys Phe Leu Pro Ser Asn Thr Val

50

55

60

aga acg tgaagtcattg agctgctgct aaggcatgtg gcaaccttga agagaagggtc 414

Arg Thr

aagagctacc agccacccaaa agaattgccag cacttcctgt gtcttttgctt tggattcatg 474

agaaatatac gttcctatatt gcttcaaaaa aaaaaaaaaa tgcgaaa 521

<210> 59

<211> 516

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 83..244

<220>

<221> sig_peptide

<222> 83..139

<223> Von Heijne matrix

score 4.24406450728052

seq LLILSKVLDDTFQ/SV

<220>

<221> polyA_site

<222> 501..516

<400> 59

actgctatag cgatataact cagatgttct caggacagca agttttacatg tgaatagaga 60

gggagggacg tgtttaggaag ac atg aaa ttc aat ttt gta ctt ttg ata ctt 112

Met Lys Phe Asn Phe Val Leu Leu Ile Leu

-15

-10

tcc aaa gtt cta gat gac act ttc caa agt gtc aag aaa tgg tta aat 160

Ser Lys Val Leu Asp Asp Thr Phe Gln Ser Val Lys Lys Trp Leu Asn

50

```

      -5              1              5
tat ttt cag ttt act tta aga aat ggt tta atg tgg cca ggt gcg gtg      208
Tyr Phe Gln Phe Thr Leu Arg Asn Gly Leu Met Trp Pro Gly Ala Val
      10              15              20
gct cat gcc tgt aat ccc agc act ggc tca cgc ctg taatcccagc      254
Ala His Ala Cys Asn Pro Ser Thr Gly Ser Arg Leu
      25              30              35
actttgggag tccgaggtgg gcagatcatg aggtcaggag atcgagacca tcctggctaa      314
catggtgaaa ccccgctctct actaaaaata caaaaaaatt agccaggcat ggtggcgggc      374
gcctgtagtc ccagctactc gggaggctga ggcaggagaa tggcgtgaac ctgggaggcg      434
gagcttacag tgagccgaga tcgcgccact gcactccagc ctgggcgaca gagcgagact      494
ctgtttcaaaa aaaaaaaaaa aa      516

```

<210> 60
 <211> 517
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 83..244

<220>
 <221> sig_peptide
 <222> 83..139
 <223> Von Heijne matrix
 score 4.24406450728052
 LLILSKVLDDTFQ/SV

<220>
 <221> polyA_site
 <222> 502..517

```

<400> 60
actgctatag cgatataact cagatgttct caggacagca agtttacatg tgaatagaga      60
gggaggggacg tgttaggaag ac atg aaa ttc aat ttt gta ctt ttg ata ctt      112
                        Met Lys Phe Asn Phe Val Leu Leu Ile Leu
                        -15              -10
tcc aaa gtt cta gat gac act ttc caa agt gtc aag aaa tgg tta aat      160
Ser Lys Val Leu Asp Asp Thr Phe Gln Ser Val Lys Lys Trp Leu Asn
                        -5              1              5
tat ttt cag ttt act tta aga aat ggt tta atg tgg cca ggt gcg gtg      208
Tyr Phe Gln Phe Thr Leu Arg Asn Gly Leu Met Trp Pro Gly Ala Val
      10              15              20
gct cat gcc tgt aat ccc agc act ggc tca cgc ctg taatcccagc      254
Ala His Ala Cys Asn Pro Ser Thr Gly Ser Arg Leu
      25              30              35
actttgggag tccgaggtgg gcagatcatg aggtcaggag atcgagacca tcctggctaa      314
catggtgaaa ccccgtytct actaaaaata caaaaaaatt agccaggcat ggtggcgggc      374
gcctgtagtc ccagctactc gggaggctga ggcaggagaa tggcgtgaac ctgggaggcg      434
gagcttacag tgagccgaga tcgcgccact gcactccagc ctgggcgaca gagcgagact      494
ctgtttcaaaa aaaaaaaaaa aaa      517

```

<210> 61
 <211> 518
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 84..245

<220>

<221> sig_peptide
 <222> 84..140
 <223> Von Heijne matrix
 score 4.24406450728052
 LLILSKVLDDTFQ/SV

<220>
 <221> polyA_site
 <222> 503..518

<400> 61
 aactgctata gcgatataac tcagatgttc tcaggacagc aagtttacat gtgaatagag 60
 agggagggac gtgttaggaa gac atg aaa ttc aat ttt gta ctt ttg ata ctt 113
 Met Lys Phe Asn Phe Val Leu Leu Ile Leu
 -15 -10
 tcc aaa gtt cta gat gac act ttc caa agt gtc aag aaa tgg tta aat 161
 Ser Lys Val Leu Asp Asp Thr Phe Gln Ser Val Lys Lys Trp Leu Asn
 -5 1 5
 tat ttt cag ttt act tta aga aat ggt tta atg tgg cca ggt gcg gtg 209
 Tyr Phe Gln Phe Thr Leu Arg Asn Gly Leu Met Trp Pro Gly Ala Val
 10 15 20
 gct cat gcc tgt aat ccc agc act ggc tca cgc ctg taatcccagc 255
 Ala His Ala Cys Asn Pro Ser Thr Gly Ser Arg Leu
 25 30 35
 actttgggag tccgaggtgg gcagatcatg aggtcaggag atcgagacca tcctgggctaa 315
 catggtgaaa ccccgctctct actaaaaata caaaaaaatt agccaggcat ggtggcgggc 375
 gcctgtagtc ccagctactc gggaggctga ggcaggagaa tggcgtgaac ctgggaggcg 435
 gagcttacag tgagccgaga tcgcgccact gcactccagc ctgggcgaca gagcgagact 495
 ctgtttcaaa aaaaaaaaaa aaa 518

<210> 62
 <211> 548
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 57..329

<220>
 <221> sig_peptide
 <222> 57..113
 <223> Von Heijne matrix
 score 4.2427958051416
 seq FVLLLSEIVSISA/LS

<220>
 <221> polyA_site
 <222> 533..548

<220>
 <221> polyA_signal
 <222> 505..510

<400> 62
 agttgtcttt ggtagtcttt ttgcactaac ttcaggaacc agtcatgat ctcagg atg 59
 Met
 tat gga aaa ata atc ttt gta tta cta ttg tca gaa att gtg agc ata 107
 Tyr Gly Lys Ile Ile Phe Val Leu Leu Leu Ser Glu Ile Val Ser Ile
 -15 -10 -5
 tca gca tta agt acc act gag gtg gca atg cac act tca acc tct tct 155
 Ser Ala Leu Ser Thr Thr Glu Val Ala Met His Thr Ser Thr Ser Ser
 1 5 10

52

[illegible]

```
<210> 63
<211> 722
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 331..672
```

```
<220>
<221> sig_peptide
<222> 331..426
<223> Von Heijne matrix
score 4.21873806574589
seq LLGSLGFVFSLTA/FN
```

```
<220>
<221> polyA_site
<222> 707..722
```

```
<220>  
<221> polyA_signal  
<222> 688..693
```

<400>	63					
t t t t t c c t c g	t c c t c t c c c g	a c a g a g c t g a	c g t g t c c t g g	g t t c c a c c g g	g a g c g g g c a t	60
t t c c a c c g g a	c g g g a g g g t t	c g g g g t g t c c	g g g g c t g g g g	a a t a c g t a g g	g g t t g c c g c g	120
c g g t g t g g g g	a g t t g g g g c g	t g t g g c t g c a	g t c c c g g g a g	t t c t t g g a g g	g g g t c g g c c c	180
a c c g a g e t t c	e g g a c c g g c t	g a t c t g c c c g	t a g c t t g c c g	g a g g g a g g g c	g g a g c c t g a c	240
t e t c c g t e c c	t t c t c c c a t c	c c c t c c a g t g	g t g g g t a c g g	g c a c t c g c t	g g c g c t c t c c	300
t c c c t c c t g t	c c c t g c t g c t	c t t t g c t g g g	a t g c a g a t g	t a c a g c c g t	c a g c t g	354
			M e t G l n M e t	T y r S e r A r g	G l n L e u	
			-30		-25	
g c c t c c a c c	g a g t g g c t c	a c c a t c c a g	g g c g g c c t g	c t t g g t t c g	g g t	402
A l a S e r T h r	G l u T r p L e u	T h r I l e G l n	G l y G l y L e u	L e u L e u G l y	S e r G l y	
	-20		-15		-10	
c t c t t c g t g	t t c t c g c t c	a c t g c c t t c	a a t a a t c t g	g a g a a t c t t	g t c	450
L e u P h e V a l	P h e S e r L e u	T h r A l a P h e	A s n A s n L e u	G l u A s n L e u	V a l	
	-5	1		5		
t t t g g c a a a	g g a t t c c a a	g c a a a g a t c	t t c c c t g a g	a t t c t c c t g	t g c	498
P h e G l y L y s	G l y P h e G l n	A l a L y s I l e	P h e P r o G l u	I l e L e u L e u	C y s	
	10	15	20			
c t c c t g t t g	g c t c t c t t t	g c a t c t g g c	c t c a t c c a c	c g a g t c t g t	g t c	546
L e u L e u L e u	A l a L e u P h e	A l a S e r G l y	L e u I l e H i s	A r g V a l C y s	V a l	
	25	30	35		40	
a c c a c c t a c	t t c a t c t t c	t c c a t a q t t	q q t c t g t a c	t a c a t c a a c	a a q	594

53

```

Thr Thr Cys Phe Ile Phe Ser Met Val Gly Leu Tyr Tyr Ile Asn Lys
      45      50      55
aty tcc tcc acc ctg tac cag gca gca gct cca gtc ctc aca cca gcc      642
Ile Ser Ser Thr Leu Tyr Gln Ala Ala Ala Pro Val Leu Thr Pro Ala
      60      65      70
aag gtc aca ggc aag agc aag aag aga aac tgaccctgaa tgttcaataa      692
Lys Val Thr Gly Lys Ser Lys Lys Arg Asn
      75      80
agttgattct ttgcaaaaaa aaaaaaaaaa      722

```

<210> 64
 <211> 479
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 96..320

<220>
 <221> sig_peptide
 <222> 96..200
 <223> Von Heijne matrix
 score 4.14350049367799
 seq MAVAFVLSLGVA/LY

<220>
 <221> polyA_site
 <222> 464..479

<220>
 <221> polyA_signal
 <222> 437..442

```

<400> 64
attctgcgcc tgcgcgcgtc tacagcacgg ttcgtttttc ctttagtcag gaaggacgtt      60
ggtgttgagc atacgtatca aggacagtaa ytacc atg gct ccc gaa gtt ttg      113
                               Met Ala Pro Glu Val Leu
                               -35      -30
cca aaa cct cgg atg cgt ggc ctt ctg gcc agg cgt ctg cga aat cat      161
Pro Lys Pro Arg Met Arg Gly Leu Leu Ala Arg Arg Leu Arg Asn His
      -25      -20      -15
atg gct gta gca ttc gtg cta tcc ctg ggg gtt gca gct ttg tat aag      209
Met Ala Val Ala Phe Val Leu Ser Leu Gly Val Ala Ala Leu Tyr Lys
      -10      -5      1
ttt cgt gtg gct gat caa aga aag aac gca tac gca gat ttc tac aga      257
Phe Arg Val Ala Asp Gln Arg Lys Asn Ala Tyr Ala Asp Phe Tyr Arg
      5      10      15
aac tac gat gtc atg aaa gat ttt gag gag atg agg aag gct ggt atc      305
Asn Tyr Asp Val Met Lys Asp Phe Glu Glu Met Arg Lys Ala Gly Ile
      20      25      30      35
ttt cag agt gta aag taatcttgga atataaagaa tttcttcagg ttgaattacc      360
Phe Gln Ser Val Lys
      40
tagaagtttg tcaactgactt gtgttcctga actatgacac atgaatatgt gggctaagaa      420
atagttcctc ttgataaata aacaattaac aaatactttg gacaaaaaaaa aaaaaaaaaa      479

```

<210> 65
 <211> 1627
 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS
<222> 90..413

<220>
<221> sig_peptide
<222> 90..209
<223> Von Heijne matrix
score 4.06476262296284
seq VTAPSMVAPVTFA/SI

<220>
<221> polyA_site
<222> 1612..1627

<220>
<221> polyA_signal
<222> 1577..1582

<400> 65
aatctctgcg aaagcttcca ggttcgctca tcgtaggtca ctggaaaaaa ttgtagcatt 60
gctaaagagt atttcagaga aaattttgc atg caa aaa ttt agg aag atg agt 113
Met Gln Lys Phe Arg Lys Met Ser
-40 -35
gaa act cat cat tct gta atc tct gtg aaa gct tcc agt ccc tgg cta 161
Glu Thr His His Ser Val Ile Ser Val Lys Ala Ser Ser Pro Trp Leu
-30 -25 -20
tct tct tca gtg act gct cca tcc atg gta gcc cca gtc act ttt gca 209
Ser Ser Ser Val Thr Ala Pro Ser Met Val Ala Pro Val Thr Phe Ala
-15 -10 -5
tct att gta gaa gaa gaa cta caa caa gaa gca gct ctt att aga agt 257
Ser Ile Val Glu Glu Glu Leu Gln Gln Glu Ala Ala Leu Ile Arg Ser
1 5 10 15
cga gaa aaa ccg ttg gct ctg att cag att gag gag cat gcc ata caa 305
Arg Glu Lys Pro Leu Ala Leu Ile Gln Ile Glu Glu His Ala Ile Gln
20 25 30
gat tta ttg gtt ttc tat gag gca ttt ggc aac cct gaa gag ttt gtc 353
Asp Leu Leu Val Phe Tyr Glu Ala Phe Gly Asn Pro Glu Glu Phe Val
35 40 45
att gtt gaa agg aca ccg cag gga cca ctg gca gta cct atg tgg aat 401
Ile Val Glu Arg Thr Pro Gln Gly Pro Leu Ala Val Pro Met Trp Asn
50 55 60
aag cat gga tgc tagttcactg tggagttgag atgcatttta cataattatg 453
Lys His Gly Cys
65
agtttggttca tataaagaaa agctgtggaa aagagtctta gagattttgt aatatcattc 513
taaataagatt aagaaaagat ataatttctt tactgcagtt aaatcatata atgtttgtat 573
gattaaaaat aaattttctca gaattgtgat ttttagtaact ttatataaaa tgtgtgagac 633
aaaaacttat taagggttaaa tagaattggt tcttstgaat ratctaacaa aggaraatat 693
aagtgattga atcataagat ataagggggg taaagtatta aaaataactt ttttgtttga 753
taacttgaga atttagaaga ttttgccaag tatgtgttgt tgcttgactt cttaaatatg 813
gcattgatga atttaaagta ggagcatcag ttattacttc tgattcatta atggccagaa 873
ttttgtgttt ggtgtaatat ttgtgtcacc attcttggtt ctttttaaaa atcaggctaa 933
tcatgtggtc catgtctctt caaagcttga cctgcacaaa tgccatattt cyatttggac 993
cacatattct ccattttgca ttgagcagta gagtacagt gaaagggaa aagaatactg 1053
attattctga acagtttagt cccaagagaa tagcgtttta aaaaagaaaa acaagatttg 1113
gagtcattgt gggttatttt tgggtgggatg gaggatctta aaaatgccta attgtgagag 1173
aatcaattgc tgaaagtgtt aaaatttctg aaaataaatg cttaattaca tatacaggaa 1233
ttaaatagtt tggaagaggg ttggattatc attaccttta caatactgta taatcagaag 1293
ttctctgaac ctcaattgta tatctagaca taaaaattgt tttctgtata ggatgttgtt 1353
tggtttgttt ctgagtgttt aaattttgca aaaacaaatg ttaaatttgt gcttcagtac 1413
ctagataaat tggaaagggt aatgttctag tttctggaag gtaagcctgg gagacacata 1473
agcaattcac tgctataatt tagttgatgt aaaatgacgg aaactggctc aatatgtcag 1533
gtttaactct gcccaaaagc agcagacatg taagcagatg tgcaataaaa aatgatcttg 1593

atccatttca aaaaaaagaa aaaaaaaaaa aaaa

1627

<210> 66

<211> 555

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 9..389

<220>

<221> sig_peptide

<222> 9..149

<223> Von Heijne matrix
 score 3.92635967755835
 seq LSSDLQQVGGASA/RI

<220>

<221> polyA_site

<222> 540..555

<220>

<221> polyA_signal

<222> 512..517

<400> 66

tgaggagtg atg ttc acg cct ctg aca gtg aaa tac gcg tac tac gac act	50
Met Phe Thr Pro Leu Thr Val Lys Tyr Ala Tyr Tyr Asp Thr	
-45 -40 -35	
gaa cgc atc gga gtt gac ctg atc atg aag acc tgc ttt agc ccc aac	98
Glu Arg Ile Gly Val Asp Leu Ile Met Lys Thr Cys Phe Ser Pro Asn	
-30 -25 -20	
aga gtg att gga ctc tca agt gac ttg cag caa gta gga ggg gca tca	146
Arg Val Ile Gly Leu Ser Ser Asp Leu Gln Gln Val Gly Gly Ala Ser	
-15 -10 -5	
gct cgc atc cag gat gcc ctg agt aca gtg ttg caa tat gca gag gat	194
Ala Arg Ile Gln Asp Ala Leu Ser Thr Val Leu Gln Tyr Ala Glu Asp	
1 5 10 15	
gta ctg tct gga aag gtg tca gct gac aat act gtg ggc cgc ttc ctg	242
Val Leu Ser Gly Lys Val Ser Ala Asp Asn Thr Val Gly Arg Phe Leu	
20 25 30	
atg agc ctg gtt aac caa gta ccg aaa ata gtt ccc gat gac ttt gag	290
Met Ser Leu Val Asn Gln Val Pro Lys Ile Val Pro Asp Asp Phe Glu	
35 40 45	
acc atg ctc aac agc aac atc aat gac ctt ttg atg gtg acc tac ctg	338
Thr Met Leu Asn Ser Asn Ile Asn Asp Leu Leu Met Val Thr Tyr Leu	
50 55 60	
gcc aac ctc aca cag tca cag att gca ctc aat gaa aaa ctt gta aac	386
Ala Asn Leu Thr Gln Ser Gln Ile Ala Leu Asn Glu Lys Leu Val Asn	
65 70 75	
ctg tgaatggacc ccaagcagta cacttgctgg tctaggtatt aaccccagga	439
Leu	
80	
ctcagaagtg aaggagaaat ggggttttttg tgggtcttgag tcacactgag atagtcagtt	499
gtgtgtgact ctaataaacg gagcctacct tttgtaaatt aaaaaaaaaa aaaaaa	555

<210> 67

<211> 422

<212> DNA

<213> Homo sapiens

<220>

<221> CDS
<222> 24..278

<220>
<221> sig_peptide
<222> 24..119
<223> Von Heijne matrix
score 3.85301646018147
seq LAHIQALISGIEA/QL

<220>
<221> polyA_site
<222> 407..422

<220>
<221> polyA_signal
<222> 384..389

<400> 67
ttgagctgca gtcacagctg agc atg aaa gct gcc ttg gaa gac aca ctg gca 53
Met Lys Ala Ala Leu Glu Asp Thr Leu Ala
-30 -25
gaa acg gag gcg cgc ttt gga gcc cag ctg gcg cat atc cag gcg ctg 101
Glu Thr Glu Ala Arg Phe Gly Ala Gln Leu Ala His Ile Gln Ala Leu
-20 -15 -10
atc agc ggt att gaa gcc cag ctg ggc gat gtg cga gct gat agt gag 149
Ile Ser Gly Ile Glu Ala Gln Leu Gly Asp Val Arg Ala Asp Ser Glu
-5 1 5 10
cgg cag aat cag gag tac cag cgg ctc atg gac atc aag tcg cgg ctg 197
Arg Gln Asn Gln Glu Tyr Gln Arg Leu Met Asp Ile Lys Ser Arg Leu
15 20 25
gag cag gag att gcc acc tac cgc agc ctg ctc gag gga cag gaa gat 245
Glu Gln Glu Ile Ala Thr Tyr Arg Ser Leu Leu Glu Gly Gln Glu Asp
30 35 40
cac tac aac aat ttg tct gcc tcc aag gtc ctc tgaggcagca ggctctgggg 298
His Tyr Asn Asn Leu Ser Ala Ser Lys Val Leu
45 50
cttctgctgt ccttttgagg gtgtcttctg ggtagaggga tgggaaggaa gggaccctta 358
cccccggtc ttctctgac ctgccataa aaatttatgg tccaaggcaa aaaaaaaaaa 418
aaaa 422

<210> 68
<211> 630
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 232..534

<220>
<221> sig_peptide
<222> 232..306
<223> Von Heijne matrix
score 3.78142478107177
seq AKTCLVLC SRVLS/VI

<220>
<221> polyA_site
<222> 615..630

<220>
<221> polyA_signal

<222> 595..600

<400> 68

```

tatcactgtt acgaaccaag gatttacaga tcactggcaa aaattctgag aactttcaca      60
ccagtatact gtccaagccc attaagtggc atcacacctc tcttttatgt agctcagaca      120
agacagtcta atatcttcaa aatactactg caatatggaa tcttagaaaag agaaaaaaac      180
cctatcaaca ttgtcttaac aatagtactc tacccttcga gagtaagagt a atg gtt      237
                                         Met Val
                                         -25
gat cgt gaa ttg gct gac atc cat gaa gat gcc aaa aca tgt ttg gta      285
Asp Arg Glu Leu Ala Asp Ile His Glu Asp Ala Lys Thr Cys Leu Val
          -20                      -15                      -10
cta tgt tcc aga gtg ctt tct gtc att tca gtc aag gaa ata aag aca      333
Leu Cys Ser Arg Val Leu Ser Val Ile Ser Val Lys Glu Ile Lys Thr
          -5                      1                      5
cag ctg agt tta gga aga cat cca att att tca aat tgg ttt gat tac      381
Gln Leu Ser Leu Gly Arg His Pro Ile Ile Ser Asn Trp Phe Asp Tyr
10                      15                      20                      25
att cct tca aca aga tac aaa gat cca tgt gaa cta tta cat ctt tgc      429
Ile Pro Ser Thr Arg Tyr Lys Asp Pro Cys Glu Leu Leu His Leu Cys
          30                      35                      40
aga cta acc atc agg aat caa cta tta acc aac aat atg ctc cca gat      477
Arg Leu Thr Ile Arg Asn Gln Leu Leu Thr Asn Asn Met Leu Pro Asp
          45                      50                      55
gga ata ttt tca ctt cta att cct gct cgt cta caa aac tat ctg aat      525
Gly Ile Phe Ser Leu Leu Ile Pro Ala Arg Leu Gln Asn Tyr Leu Asn
          60                      65                      70
tta gaa atc taacatacgt cagtgtccta agttccttaa caatgcttac      574
Leu Glu Ile
          75
caatgtatgg cttagaagtt aataaaaatt cacttcatgc aaaaaaaaaa aaaaat      630

```

<210> 69

<211> 730

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 80..283

<220>

<221> sig_peptide

<222> 80..220

<223> Von Heijne matrix

score 3.72491712201476

seq PLLSLHSRGGSSS/ES

<220>

<221> polyA_site

<222> 715..730

<220>

<221> polyA_signal

<222> 697..702

<400> 69

```

attttgcgaa cggcgagcag cggcggcggc gcggagagac gcagcggagg ttttcctggt      60
ttcggacccc agcggccgg atg gtg aaa tcc tcc ctg cag cgg atc ctc aat      112
                      Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn
                      -45                      -40
agc cac tgc ttc gcc aga gag aag gaa ggg gat aaa ccc agc gcc acc      160
Ser His Cys Phe Ala Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr

```

```

      -35              -30              -25
atc cac gcc agc cgc acc atg ccg ctc cta agc ctg cac agc cgc ggc      208
Ile His Ala Ser Arg Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly
-20              -15              -10              -5
ggc agc agc agt gag agt tcc agg gtc tcc ctc cac tgc tgt agt aac      256
Gly Ser Ser Ser Glu Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn
              1              5              10
ccg ggt ccg ggg cct cgg tgg tgc tcc tgatgcccct caccaccccc      303
Pro Gly Pro Gly Pro Arg Trp Cys Ser
      15              20
tgaagatccc aggtgggcga ggaggaggag tagggccgcc tcggggctgg gcatccggcc      363
cctggggcca ccccttgtca gccgggtggg taggaaccgt agactcgctc atctcgctg      423
ggtttgtccg catgttgtaa tcgtgcaa ataacgctcac tccgaattag cggtgtattt      483
cttgaagttt aatatttgtt ttgtgatact gaagtatttg ctttaattct aaataaaaaa      543
ttatatttta cttttttatt gctgggttaa gatgattcag attatccttg tactttgagg      603
agaagtttct tattttggagt cttttggaaa cagtcttagt cttttaactt ggaaagatga      663
ggtattaatc ccctccattg ctctccaaaa gccaataaag tgattacacc caaaaaaaaaa      723
aaaaaaa      730

```

<210> 70
 <211> 730
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 80..283

<220>
 <221> sig_peptide
 <222> 80..220
 <223> Von Heijne matrix
 score 3.72491712201476
 seq PLLSLHSRGGSSS/ES

<220>
 <221> polyA_site
 <222> 715..730

<220>
 <221> polyA_signal
 <222> 697..702

```

<400> 70
attttgcgaa cggcgagcag cggcggcggc gcggagagac gcagcggagg ttttcctggt      60
ttcggacccc agcggccgg atg gtg aaa tcc tcc ctg cag cgg atc ctc aat      112
              Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn
              -45              -40
agc cac tgc ttc gcc aga gag aag gaa ggg gat aaa ccc agc gcc acc      160
Ser His Cys Phe Ala Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr
      -35              -30              -25
atc cac gcc agc cgc acc atg ccg ctc cta agc ctg cac agc cgc ggc      208
Ile His Ala Ser Arg Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly
-20              -15              -10              -5
ggc agc agc agt gag agt tcc agg gtc tcc ctc cac tgc tgt agt aac      256
Gly Ser Ser Ser Glu Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn
              1              5              10
ccg ggt ccg ggg cct cgg tgg tgc tcc tgatgcccct caccaccccc      303
Pro Gly Pro Gly Pro Arg Trp Cys Ser
      15              20
tgaagatccc aggtgggcga ggaggaggag tagggccgcc tcggggctgg gcatccggcc      363
cctggggcca ccccttgtca gccgggtggg taggaaccgt agactcgctc atctcgctg      423
ggtttgtccg catgttgtaa tcgtgcaa ataacgctcac tccgaattag cggtgtattt      483

```

```

cttgaagttt aatattgtgt ttgtgatact gaagtatttg ctttaattct aaataaaaat 543
ttatatattta cttttttatt gctgggttaa gatgattcag attatccttg tactttgagg 603
agaagtttct tatttgaggt cttttggaaa cagtcttagt cttttaactt ggaaagatga 663
ggtattaatc ccctccattg ctctccaaaa gccataaag tgattacacc caaaaaaaaaa 723
aaaaaaaaa 730

```

```

<210> 71
<211> 863
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> CDS
<222> 80..283

```

```

<220>
<221> sig_peptide
<222> 80..220
<223> Von Heijne matrix
      score 3.72491712201476
      PLLSLHSRGGSSS/ES

```

```

<220>
<221> polyA_site
<222> 844..863

```

```

<220>
<221> polyA_signal
<222> 827..832

```

```

<400> 71
attttgcgaa cggcgagcag cggcggcggc gcggagagac gcagcggagg ttttcctggt 60
ttcggacccc agcggccgg atg gtg aaa tcc tcc ctg cag cgg atc ctc aat 112
                Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn
                -45                                -40
agc cac tgc ttc gcc aga gag aag gaa ggg gat aaa ccc agc gcc acc 160
Ser His Cys Phe Ala Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr
-35                                -30                                -25
atc cac gcc agc cgc acc atg ccg ctc cta agc ctg cac agc cgc ggc 208
Ile His Ala Ser Arg Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly
-20                                -15                                -10                                -5
ggc agc agc agt gag agt tcc agg gtc tcc ctc cac tgc tgt agt aac 256
Gly Ser Ser Ser Glu Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn
                1                                5                                10
ccg ggt ccg ggg cct cgg tgg tgc tcc tgatgccct caccacccc 303
Pro Gly Pro Gly Pro Arg Trp Cys Ser
                15                                20
tgaagatccc aggtgggoga gggaatagtc agagggatca caatctttca gctaacttat 363
tctactccga tgatcggtctg aatgtaacag aggaactaac gtccaacgac aagacgagga 423
ttctcaacgt ctacgcctct acatcgagat cccgggcggc gcgctgcccg aggggagcaa 483
ggacagcttt gcagttctcc tggagttcgc tgaggagcag ctgcgagccg accatgtctt 543
catttgcttc cacaagaacc gcgaggacag agccgccttg ctccgaacct tcagcttttt 603
gggctttgag attgtgagac cggggcatcc ccttgtcccc aagagaccog acgcttgctt 663
catggcctac acgttcgaga gagagtcttc gggagaggag gaggagtagg gccgcctcgg 723
ggctgggcat ccggcccctg gggccacccc ttgtcagccg ggtgggtagg aaccgtagac 783
tcgctcatct cgcttgggtt tgtccgcatg ttgtaatcgt gcaaataaac gctcactccg 843
aattaaaaaa aaaaaaaaaa 863

```

```

<210> 72
<211> 730
<212> DNA
<213> Homo sapiens

```

<220>

<221> CDS

<222> 80..283

<220>

<221> sig_peptide

<222> 80..220

<223> Von Heijne matrix

score 3.72491712201476

PLLSSLHSRGGSSS/ES

<220>

<221> polyA_site

<222> 715..730

<220>

<221> polyA_signal

<222> 697..702

<400> 72

```

attttgcgaa cggcgagcag cggcggcgcc gcggagagac gcagcggagg ttttcctggt      60
ttcggacccc agcggcccg atg gtg aaa tcc tcc ctg cag cgg atc ctc aat      112
                Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn
                -45                                -40

agc cac tgc ttc gcc aga gag aag gaa ggg gat aaa ccc agc gcc acc      160
Ser His Cys Phe Ala Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr
   -35                -30                -25

atc cac gcc agc cgc acc atg ccg ctc cta agc ctg cac agc cgc ggc      208
Ile His Ala Ser Arg Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly
-20                -15                -10                -5

ggc agc agc agt gag agt tcc agg gtc tcc ctc cay tgy tgt agt aas      256
Gly Ser Ser Ser Glu Ser Ser Arg Val Ser Leu His Cys Cys Ser Xaa
                1                5                10

ccg ggt ccg ggg cct cgg tgg tgc tcc tgatgccct caccaccccc      303
Pro Gly Pro Gly Pro Arg Trp Cys Ser
   15                20

tgaagatccc aggtgggcga ggaggaggag tagggccgcc tcggggctgg gcatccggcc      363
cctggggcca ccccttgta gccgggtggg taggaaccgt agactcgctc atctcgctg      423
ggtttgtccg catgttgtaa tcgtgcaaat aaacgctcac tccgaattag cggtgtattt      483
cttgaagttt aatatttgtt ttgtgatact gaagtatttg ctttaattct aaataaaaaat      543
ttatatttta cttttttatt gctggtttaa gatgattcag attatccttg tactttgagg      603
agaagtttct tatttggagt cttttggaaa cagtcttagt cttttaactt ggaaagatga      663
ggtattaatc ccctccattg ctctccaaaa gccataaag tgattacacc caaaaaaaaaa      723
aaaaaaaaa

```

<210> 73

<211> 879

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 9..395

<220>

<221> sig_peptide

<222> 9..56

<223> Von Heijne matrix

score 3.61791418904325

LLLLLRALRRGPG/PG

<220>

<221> polyA_site

<222> 864..879

<400> 73

```

aggccaac atg gcc gtg ctg ctg ctg ctg ctc cgt gcc ctc cgc cgg ggt      50
          Met Ala Val Leu Leu Leu Leu Leu Arg Ala Leu Arg Arg Gly
          -15                               -10                               -5

cca ggc ccg ggt cct cgg ccg ctg tgg ggc cca ggc ccg gcc tgg agt      98
Pro Gly Pro Gly Pro Arg Pro Leu Trp Gly Pro Gly Pro Ala Trp Ser
          1                               5                               10

cca ggg ttc ccc gcc agg ccc ggg agg ggg cgg ccg tac atg gcc agc      146
Pro Gly Phe Pro Ala Arg Pro Gly Arg Gly Arg Pro Tyr Met Ala Ser
15                               20                               25                               30

agg cct ccg ggg gac ctc gcc gag gct gga ggc cga gct ctg cag agc      194
Arg Pro Pro Gly Asp Leu Ala Glu Ala Gly Gly Arg Ala Leu Gln Ser
          35                               40                               45

tta caa ttg aga ctg cta acc cct acc ttt gaa ggg atc aac gga ttg      242
Leu Gln Leu Arg Leu Leu Thr Pro Thr Phe Glu Gly Ile Asn Gly Leu
          50                               55                               60

ttg ttg aaa caa cat tta gtt cag aat cca gtc aga ctc tgg caa ctt      290
Leu Leu Lys Gln His Leu Val Gln Asn Pro Val Arg Leu Trp Gln Leu
          65                               70                               75

tta ggt ggt act ttc twt ttt aac acc tca agg ttg aag cak aag aat      338
Leu Gly Gly Thr Phe Xaa Phe Asn Thr Ser Arg Leu Lys Xaa Lys Asn
          80                               85                               90

aag gag aag gat aag tcg aag ggg aag gcg cct gaa gag gac gaa ggt      386
Lys Glu Lys Asp Lys Ser Lys Gly Lys Ala Pro Glu Glu Asp Glu Gly
          95                               100                               105                               110

ata ttc atc tgatgttctt cagtcagtag ctgcctctgg atgtctttac      435
Ile Phe Ile

atttctgttt tcctttttagc aaggtgaaac cagtctggaa aatggggaga tgggccgggt      495
gcagtggctc acacttgtaa tcgaaacgct ttgggaggcc caggtggaag gatcacttga      555
ggcctatacc acatagcgag accctgtctc actgcaaatt aaaaggctgg gcgtggtggc      615
tcacacctgt aatcccagca ctttgggagg ctgaggcagg cggatcacct gcaccctggc      675
caacatggtg aaaccccgctc tttaactaaaa atagaaaatt agccggggcgt gatggcacac      735
gcctgtaatc ccagctactc gggaggctga ggcaggagaa ttgcttgaac ctgggaggtg      795
gaggttgctg tgagtggaga tcatgccatt gcactccagc ctgagcaaca agagcaaaac      855
tccatctcaa aaaaaaaaaa aaaa      879

```

<210> 74

<211> 879

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 9..395

<220>

<221> sig_peptide

<222> 9..56

<223> Von Heijne matrix

score 3.61791418904325

LLLLLRALRRGPG/PG

<220>

<221> polyA_site

<222> 864..879

<400> 74

```

aggccaac atg gcc gtg ctg ctg ctg ctg ctc cgt gcc ctc cgc cgg ggt      50
          Met Ala Val Leu Leu Leu Leu Leu Arg Ala Leu Arg Arg Gly
          -15                               -10                               -5

cca ggc ccg ggt cct cgg ccg ctg tgg ggc cca ggc ccg gcc tgg agt      98

```

62

[illegible]

```
<210> 75
<211> 464
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 16..342
```

```
<220>
<221> sig_peptide
<222> 16..114
<223> Von Heijne matrix
score 3.66786754931606
seq LAVFQMLKSMCAG/QR
```

```
<220>
<221> polyA_site
<222> 446..464
```

```
<220>
<221> polyA_signal
<222> 412..417
```

```

<400> 75
gaatgcgccct gcggtt atg gcg gcc tcc ggc gcc cca agg atc ctg gtg gac      51
                Met Ala Ala Ser Gly Ala Pro Arg Ile Leu Val Asp
                -30                                -25

ctg ctg aag ctg aac gtg gcc ccc ctc gcc gtc ttc cag atg ctc aag      99
Leu Leu Lys Leu Asn Val Ala Pro Leu Ala Val Phe Gln Met Leu Lys
   -20                                -15                                -10

tcc atg tgt gcc ggg cag agg cta gcg agc gag ccc cag gac cct gcg    147

```

63

```

Ser Met Cys Ala Gly Gln Arg Leu Ala Ser Glu Pro Gln Asp Pro Ala
-5          1          5          10
gcc gtg tct ctg ccc acg tcg agc gtg ccc gag acc cga ggg aga aac      195
Ala Val Ser Leu Pro Thr Ser Ser Val Pro Glu Thr Arg Gly Arg Asn
          15          20          25
aaa ggc agc gct gcc ctc ggg gga gca ttg gcc ctg gcg gaa cgm rrc      243
Lys Gly Ser Ala Ala Leu Gly Gly Ala Leu Ala Leu Ala Glu Arg Xaa
          30          35          40
agc cgc gaa gga tcc agc cag agg atg cca cgc cag ccc agc gct acc      291
Ser Arg Glu Gly Ser Ser Gln Arg Met Pro Arg Gln Pro Ser Ala Thr
          45          50          55
agg ctg ccc aag ggg ggc ggg cct ggg aag agc cct aca cgg ggc agc      339
Arg Leu Pro Lys Gly Gly Gly Pro Gly Lys Ser Pro Thr Arg Gly Ser
60          65          70          75
acc taggatgggg cagagacttg ttgcatcttt gtccccagca aaggctacat      392
Thr
gttacctcct tcaattgata ataaaccttt ctgagatgca gaggggtccag gtcaaaaaaa      452
aaaaaaaaaa aa      464

```

<210> 76
 <211> 624
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 59..367

<220>
 <221> sig_peptide
 <222> 59..166
 <223> Von Heijne matrix
 score 3.66643500261364
 seq RALSTFLFGSIRG/AA

<220>
 <221> polyA_site
 <222> 609..624

<220>
 <221> polyA_signal
 <222> 588..593

```

<400> 76
agtgcgcaga cgcaggggtc ggcgcgggt gagagcgtgc ggccggattc accacaac      58
atg gca aat ctt ttt ata agg aaa atg gtg aac cct ctg ctc tat ctc      106
Met Ala Asn Leu Phe Ile Arg Lys Met Val Asn Pro Leu Leu Tyr Leu
-35          -30          -25
cgt cgt cac acg gtg aag cct cga gcc ctc tcc aca ttt cta ttt gga      154
Arg Arg His Thr Val Lys Pro Arg Ala Leu Ser Thr Phe Leu Phe Gly
-20          -15          -10          -5
tcc att cga ggt gca gcc ccc gtg gct gtg gaa ccc ggg gca gca gtg      202
Ser Ile Arg Gly Ala Ala Pro Val Ala Val Glu Pro Gly Ala Ala Val
          1          5          10
cgc tca ctt ctc tca ccc ggc ctc ctg ccc cat ctg ctg cct gcg ctg      250
Arg Ser Leu Leu Ser Pro Gly Leu Leu Pro His Leu Leu Pro Ala Leu
          15          20          25
ggg ttc aaa aac aag act gtc ctt aag aag cgc tgc aag gac tgt tac      298
Gly Phe Lys Asn Lys Thr Val Leu Lys Lys Arg Cys Lys Asp Cys Tyr
30          35          40
ctg gtg aag agg cgg ggt cgg tgg tac gtc tac tgt aaa acc cat ccg      346
Leu Val Lys Arg Arg Gly Arg Trp Tyr Val Tyr Cys Lys Thr His Pro
45          50          55          60

```


agg cac aag cag aga cag atg tagacccttt ccctccagag tcacgcacat 397
 Arg His Lys Gln Arg Gln Met

65

actcgtcatc gcataccttg ggagaatggt tgtatcttat ggaaggaatt atcacatcaa 457
 ggagtcaggg gaaagtgact ggaagcaaac gccctaaaag ttacccatca cgtttcagtg 517
 taaatgagta actatagaag acattgcgtt atcttatttc maaaacgttc caactaaaaa 577
 acattttcct attaaaatag accttccgaa caaaaaaaaa aaaaaaa 624

<210> 77

<211> 1158

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 58..315

<220>

<221> sig_peptide

<222> 58..165

<223> Von Heijne matrix

score 3.68684848854945

seq SVSLSGFTVGTLS/ET

<220>

<221> polyA_site

<222> 1143..1158

<400> 77

ctgtgggaga ggccaggatg ctgcattagg cacaggataa cctgggaacc caggcac 57
 atg ggt cct gct ctc cga agt ctg caa gtc aag aag gga aca gag cac 105
 Met Gly Pro Ala Leu Arg Ser Leu Gln Val Lys Lys Gly Thr Glu His
 -35 -30 -25

gcc gac cct ctc cct ttc ccc tct gtc tct ctt agt ggc ttt aca gtg 153
 Ala Asp Pro Leu Pro Phe Pro Ser Val Ser Leu Ser Gly Phe Thr Val
 -20 -15 -10 -5

ggt acc ctg tca gaa acc agc act ggg ggc cct gcc acc ccc aca tgg 201
 Gly Thr Leu Ser Glu Thr Ser Thr Gly Gly Pro Ala Thr Pro Thr Trp
 1 5 10

aag gag tgt cct atc tgt aag gag cgc ttt cct gct gag agt gac aag 249
 Lys Glu Cys Pro Ile Cys Lys Glu Arg Phe Pro Ala Glu Ser Asp Lys
 15 20 25

gat gcc ctg gag gac cac atg gat gga cac ttc ttt ttc agc acc cag 297
 Asp Ala Leu Glu Asp His Met Asp Gly His Phe Phe Ser Thr Gln
 30 35 40

gac ccc ttc acc ttt gag tgatcttact ccctcgtaca tgcacaaata 345
 Asp Pro Phe Thr Phe Glu
 45 50

cacactcatg cacacacaca ctacacacaca tgcatacact taggtttcat gccattttc 405

tatcacactg ggctccatga tattctgttc cctaagaact gcttctgtgt gccctgtttt 465

catcccaaga tttctcactt catcctctcc tacctggctc ttttgtccca gggaggggtc 525

ctgttcggaa gcagtggctg aatttatccc ctgaaagtgg ttttggagga accgggatgg 585

aggaggcctt ccctctgtggg aatagaatcg tccactccta gccctggttg cttctgatac 645

acagccactg cacacacaca ctacacactca cactcccttg tytgatgcc caaagccaat 705

tcctggggca ccctaccctc tcttatttgg agtttccgtt ggtttacctg agttttctct 765

ggggtctgca cagaggcagc agcatggaca tcatggcctc tcaggtcctt tttggttctc 825

agtttcattg gttcctcttt ctgttcccc attgacttct gtgccccacc ctageccttt 885

ccataacctt aggtattcag tttggagggg ttttttgtat ttttgaggat tcctgtattc 945

tgtatcctct cctcgcatct cctcacatgg aaagaaataa tgtatttgtg cttctgtga 1005

ggaatggggg gaacaagtgg tcccaggtat cccattttcc aaggccccc tccctctcca 1065

ggtccccca cagcaataaa agcttcccc tgatatccat ccctttgtag tttgaacaaa 1125

tatatttata tgatatgaaa aaaaaaaaaaaa aaa 1158

<210> 78
 <211> 641
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 187..342

<220>
 <221> sig_peptide
 <222> 187..276
 <223> Von Heijne matrix
 score 3.57432920610252
 seq VKLLKLMVSSATT/ER

<220>
 <221> polyA_site
 <222> 626..641

<400> 78
 attctaaatc aataaatata gcttaccac agacttttta tgggatgtaa agtatgtatt 60
 acaacaggta aatccatttc tctgttttgg gaaagcaggt taatgttttg tttgttttgt 120
 tgctgcatat aatgttgcaac tatatatata ttctaatagca gttgggtctcc taattctaca 180
 ggataa atg tct agc tgt ggc att gtt gga tca tca gtt tcg ttt cag 228
 Met Ser Ser Cys Gly Ile Val Gly Ser Ser Val Ser Phe Gln
 -30 -25 -20
 tta gat gct gtg aaa ttg ctc ttg aaa atg gtg tcc tct gcc acc aca 276
 Leu Asp Ala Val Lys Leu Leu Leu Lys Met Val Ser Ser Ala Thr Thr
 -15 -10 -5
 gaa cgg tgt tgt aat gga agt gcc aat ttt cat aaa aac ttg tgt gca 324
 Glu Arg Cys Cys Asn Gly Ser Ala Asn Phe His Lys Asn Leu Cys Ala
 1 5 10 15
 aca ggt att aag aat ttt tgaggtcggg cgcagtagct cagcctgta 372
 Thr Gly Ile Lys Asn Phe
 20
 atcccagcac tttgggaggc ctaggcaggt ggatcacctg aggtcaggag ttccagacca 432
 gcctgaccaa catggtgaaa tcccattctcc actaaataca aaaaattagc tgggcgtggt 492
 ggcatacacc tgtaatccca gctacttggg aggctaagac aggagaattg cttgaaccca 552
 ggaggcggag gttgcagtga gccgagattg caccactaca ctccagcctg ggtgatgagc 612
 gaaactccat ctcaaaaaaaaa aaaaaaaaaa 641

<210> 79
 <211> 660
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 2..658

<220>
 <221> sig_peptide
 <222> 2..121
 <223> Von Heijne matrix
 score 6.4143157541603
 ITMSATVVDAVNA/AP

<400> 79
 t cct awt cga aaa att acg aga ttg gct tgg att ctg tcg gat gga ctt 49
 Pro Xaa Arg Lys Ile Thr Arg Leu Ala Trp Ile Leu Ser Asp Gly Leu
 -40 -35 -30 -25
 ggg gct agc tgc ggc ggg gct gga gga ggc cag ata acc atg tca gcc 97

66

Gly	Ala	Ser	Cys	Gly	Gly	Ala	Gly	Gly	Gly	Gln	Ile	Thr	Met	Ser	Ala		
				-20					-15					-10			
aca	gtt	gta	gat	gca	gtt	aat	gct	gca	ccc	cta	tcg	ggg	tcc	aaa	gaa		145
Thr	Val	Val	Asp	Ala	Val	Asn	Ala	Ala	Pro	Leu	Ser	Gly	Ser	Lys	Glu		
			-5				1					5					
atg	agt	ttg	gaa	saa	cca	aag	aag	atg	acc	aga	gag	gac	tgg	aga	aag		193
Met	Ser	Leu	Glu	Xaa	Pro	Lys	Lys	Met	Thr	Arg	Glu	Asp	Trp	Arg	Lys		
	10					15			20								
aag	aag	gag	cta	gaa	gaa	cag	cga	aaa	ttg	ggc	aat	gct	cct	gca	gaa		241
Lys	Lys	Glu	Leu	Glu	Glu	Gln	Arg	Lys	Leu	Gly	Asn	Ala	Pro	Ala	Glu		
	25				30				35						40		
gtt	gat	gaa	gaa	gga	aaa	gac	atc	aac	ccc	cat	att	cct	cag	tat	att		289
Val	Asp	Glu	Glu	Gly	Lys	Asp	Ile	Asn	Pro	His	Ile	Pro	Gln	Tyr	Ile		
				45				50						55			
tct	tca	gtg	cca	tgg	tat	att	gat	cct	tca	aaa	aga	cct	act	tta	aaa		337
Ser	Ser	Val	Pro	Trp	Tyr	Ile	Asp	Pro	Ser	Lys	Arg	Pro	Thr	Leu	Lys		
			60					65					70				
cac	cag	aga	cca	caa	cca	gaa	aaa	caa	aag	cag	ttc	agc	tca	tct	gga		385
His	Gln	Arg	Pro	Gln	Pro	Glu	Lys	Gln	Lys	Gln	Phe	Ser	Ser	Ser	Gly		
			75				80					85					
gaa	tgg	tac	aag	agg	ggt	gta	aaa	gag	aat	tcc	ata	att	act	aag	tac		433
Glu	Trp	Tyr	Lys	Arg	Gly	Val	Lys	Glu	Asn	Ser	Ile	Ile	Thr	Lys	Tyr		
	90					95					100						
cgc	aaa	gga	gca	tgt	gaa	aat	tgt	ggg	gcc	atg	aca	cac	aaa	aag	aaa		481
Arg	Lys	Gly	Ala	Cys	Glu	Asn	Cys	Gly	Ala	Met	Thr	His	Lys	Lys	Lys		
	105			110				115						120			
gac	tgc	ttt	gag	aga	cct	agg	cga	gtt	gga	gcc	aaa	ttt	aca	ggt	act		529
Asp	Cys	Phe	Glu	Arg	Pro	Arg	Arg	Val	Gly	Ala	Lys	Phe	Thr	Gly	Thr		
				125				130						135			
aat	ata	gct	cca	gat	gaa	cat	gtc	cag	cct	caa	ctg	atg	ttt	gac	tat		577
Asn	Ile	Ala	Pro	Asp	Glu	His	Val	Gln	Pro	Gln	Leu	Met	Phe	Asp	Tyr		
			140					145					150				
gat	ggg	aag	agg	gat	cgg	tgg	aat	ggc	tac	aat	cca	gaa	gaa	cac	atg		625
Asp	Gly	Lys	Arg	Asp	Arg	Trp	Asn	Gly	Tyr	Asn	Pro	Glu	Glu	His	Met		
		155				160						165					
aaa	att	gtt	gaa	gag	tat	gcc	aaa	gtt	gat	ttg	gc						660
Lys	Ile	Val	Glu	Glu	Tyr	Ala	Lys	Val	Asp	Leu							
	170					175											

<210> 80

<211> 647

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 1..645

<220>

<221> sig_peptide

<222> 1..108

<223> Von Heijne matrix
 score 5.52542399647426
 ITMSATVXDAVNA/AP

<400> 80

att	acg	aga	ttg	gct	tgg	att	ctg	tck	gat	gga	ctt	ggg	gct	agc	tgc		48
Ile	Thr	Arg	Leu	Ala	Trp	Ile	Leu	Ser	Asp	Gly	Leu	Gly	Ala	Ser	Cys		
	-35					-30					-25						
ggc	ggg	gct	gga	gga	ggc	cag	ata	acc	atg	tca	gcc	aca	gtt	gtm	gat		96
Gly	Gly	Ala	Gly	Gly	Gln	Ile	Thr	Met	Ser	Ala	Thr	Val	Val	Val	Asp		
	-20				-15				-10					-5			
gca	gtt	aat	gct	gca	ccc	cta	tcg	ggg	tcc	aaa	gaa	atg	agt	ttg	gaa		144

67

Ala	Val	Asn	Ala	Ala	Pro	Leu	Ser	Gly	Ser	Lys	Glu	Met	Ser	Leu	Glu	
				1				5					10			
gaa	cca	aag	aag	atg	acc	aga	gag	gac	tgg	aga	aag	aag	aag	gag	cta	192
Glu	Pro	Lys	Lys	Met	Thr	Arg	Glu	Asp	Trp	Arg	Lys	Lys	Lys	Glu	Leu	
		15					20				25					
gaa	gaa	cag	cga	aaa	ttg	ggc	aat	gct	cct	gca	gaa	gtt	gat	gaa	gaa	240
Glu	Glu	Gln	Arg	Lys	Leu	Gly	Asn	Ala	Pro	Ala	Glu	Val	Asp	Glu	Glu	
	30					35					40					
gga	aaa	gac	atc	aac	ccc	cat	att	cct	cag	tat	att	tct	tca	gtg	cca	288
Gly	Lys	Asp	Ile	Asn	Pro	His	Ile	Pro	Gln	Tyr	Ile	Ser	Ser	Val	Pro	
45				50						55					60	
tgg	tat	att	gat	cct	tca	aaa	aga	cct	act	tta	aaa	cac	cag	aga	cca	336
Trp	Tyr	Ile	Asp	Pro	Ser	Lys	Arg	Pro	Thr	Leu	Lys	His	Gln	Arg	Pro	
				65					70					75		
caa	cca	gaa	aaa	caa	aag	cag	ttc	agc	tca	tct	gga	gaa	tgg	tac	aag	384
Gln	Pro	Glu	Lys	Gln	Lys	Gln	Phe	Ser	Ser	Ser	Gly	Glu	Trp	Tyr	Lys	
			80					85					90			
agg	ggt	gta	aaa	gag	aat	tcc	ata	att	act	aag	tac	cgc	aaa	gga	gca	432
Arg	Gly	Val	Lys	Glu	Asn	Ser	Ile	Ile	Thr	Lys	Tyr	Arg	Lys	Gly	Ala	
		95				100					105					
tgt	gaa	aat	tgt	ggg	gcc	atg	aca	cac	aaa	aag	aaa	gac	tgc	ttt	gag	480
Cys	Glu	Asn	Cys	Gly	Ala	Met	Thr	His	Lys	Lys	Lys	Asp	Cys	Phe	Glu	
	110					115					120					
aga	cct	agg	cga	gtt	gga	gcc	aaa	ttt	aca	ggt	act	aat	ata	gct	cca	528
Arg	Pro	Arg	Arg	Val	Gly	Ala	Lys	Phe	Thr	Gly	Thr	Asn	Ile	Ala	Pro	
	125				130					135				140		
gat	gaa	cat	gtc	cag	cct	caa	ctg	atg	ttt	gac	tat	gat	ggg	aag	agg	576
Asp	Glu	His	Val	Gln	Pro	Gln	Leu	Met	Phe	Asp	Tyr	Asp	Gly	Lys	Arg	
				145					150					155		
gat	cgg	tgg	aat	ggc	tac	aat	cca	gaa	gaa	cac	atg	aaa	att	gtt	gaa	624
Asp	Arg	Trp	Asn	Gly	Tyr	Asn	Pro	Glu	Glu	His	Met	Lys	Ile	Val	Glu	
			160					165					170			
gag	tat	gcc	aaa	gtt	gat	ttg	gc									647
Glu	Tyr	Ala	Lys	Val	Asp	Leu										
			175													

<210> 81
 <211> 643
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 1..642

<220>
 <221> sig_peptide
 <222> 1..108
 <223> Von Heijne matrix
 score 6.4143157541603
 seq ITMSATVVDAVNA/AP

<400>	81															
att	acg	aga	ttg	gct	tgg	att	ctg	tcg	gat	gga	ctt	ggg	gct	agc	tgc	48
Ile	Thr	Arg	Leu	Ala	Trp	Ile	Leu	Ser	Asp	Gly	Leu	Gly	Ala	Ser	Cys	
	-35					-30					-25					
ggc	ggg	gct	gga	gga	ggc	cag	ata	acc	atg	tca	gcc	aca	gtt	gta	gat	96
Gly	Gly	Ala	Gly	Gly	Gly	Gln	Ile	Thr	Met	Ser	Ala	Thr	Val	Val	Asp	
	-20				-15				-10					-5		
gca	gtt	aat	gct	gca	ccc	cta	tcg	ggg	tcc	aaa	gaa	atg	agt	ttg	gaa	144
Ala	Val	Asn	Ala	Ala	Pro	Leu	Ser	Gly	Ser	Lys	Glu	Met	Ser	Leu	Glu	
			1					5					10			
gaa	cca	aag	aag	atg	acc	aga	gag	gac	tgg	aga	aag	aag	aag	gag	cta	192

68

[illegible]

```
<210> 82
<211> 1200
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 29..469
```

```
<220>
<221> sig_peptide
<222> 29..73
<223> Von Heijne matrix
score 4.5958301326157
seq VLVSXCSARLLQQ/EE
```

```
<220>
<221> polyA_site
<222> 1185..1200
```

```
<400> 82
acttggcgag tgagacgctg atgggagg atg gac gta ctg gtg tct rag tgc      52
                        Met Asp Val Leu Val Ser Xaa Cys
                        -15                      -10

tcc gcg cgg ctg ctg cag cag gaa gaa gag att aaa tct ctg act gct      100
Ser Ala Arg Leu Leu Gln Gln Glu Glu Glu Ile Lys Ser Leu Thr Ala
          -5                          1              5

gaa att gac cgg ttg aaa aac tgt ggc tgt tta gga gct tct cca aat      148
Glu Ile Asp Arg Leu Lys Asn Cys Gly Cys Leu Gly Ala Ser Pro Asn
10                15                    20                25
```

69

```

ttg gag cag tta caa gaa gaa aat tta aaa tta aag tat cga ctg aat      196
Leu Glu Gln Leu Gln Glu Glu Asn Leu Lys Leu Lys Tyr Arg Leu Asn
                               30                               35                               40
att ctt cga aag agt ctt cag gca gaa agg aac aaa cca act aaa aat      244
Ile Leu Arg Lys Ser Leu Gln Ala Glu Arg Asn Lys Pro Thr Lys Asn
                               45                               50                               55
atg att aac att att agc cgc cta caa gag gtc ttt ggt cat gca att      292
Met Ile Asn Ile Ile Ser Arg Leu Gln Glu Val Phe Gly His Ala Ile
                               60                               65                               70
aag gct gca tat cca gat ttg gaa aat cct cct ctg cta gtg aca cca      340
Lys Ala Ala Tyr Pro Asp Leu Glu Asn Pro Pro Leu Leu Val Thr Pro
                               75                               80                               85
agt cag cag gcc aag ttt ggg gac tat cag tgt aat agt gct atg ggt      388
Ser Gln Gln Ala Lys Phe Gly Asp Tyr Gln Cys Asn Ser Ala Met Gly
90                               95                               100                               105
att tct cag gtg atg tat tgt cat gac tct tgg ctg ttt gat ttt ttt      436
Ile Ser Gln Val Met Tyr Cys His Asp Ser Trp Leu Phe Asp Phe Phe
                               110                               115                               120
aag tat tat tat cat cat tgc cat tta cag aaa taatactatt acaagttgta      489
Lys Tyr Tyr Tyr His His Cys His Leu Gln Lys
                               125                               130
tccttagtga aaaggacatt tgccacagtt tgaaaaactt gagaaaggag ttggggggggt      549
atatgtttta acttttttag gcacaatttt taaggtttgg ttaaatttta tatgtattct      609
caatatttaa gggcaatcat tgggtactott ttgttttaggt atttccctcc tgctgtgtcc      669
aggaktgctg tgtggtggtr atgagtgtctg ggaggtgaaa aattaaaata agccatttac      729
cagtcagcat cccaattaaa tatttgatgt aactgtgatc tttgagccag gcttatatat      789
tcattttcaa gcagaggagt tccccatttt aaatagaggc attgtctgat gtgtttatgg      849
ttaactgcat ctggcttggg tctttctgtt ttcctttctt tgctgaattg gaaggggtta      909
ctctgaagag tccaggtcct acagtgtggt ttatttctca agtgtgaata ttgcacacct      969
tcatggcttg aaaattagaa atgtaaatat gctggaacca gacttcataa ctgtagactt     1029
tgttatgtca tcattaaaaa gttcgagggtg ttgtaatccc agctactagg gaggctgagg     1089
caggagaatt gcttgaaccc aggggcggag gttacagtga gccgagatca caccactgca     1149
ctccagcctg ggcaacagag cgagactcca tctctaaaaa aaaaaaaaaa a              1200

```

<210> 83

<211> 351

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 1..351

<220>

<221> sig_peptide

<222> 1..129

<223> Von Heijne matrix
score 4.10872570449087
LGQVAAAAANA/AS

<220>

<221> polyA_site

<222> 336..351

<220>

<221> polyA_signal

<222> 306..311

<400> 83

```

ctt tct cca gct caa gcg tct ggt att tgg tgt agg gac agc gct ttt      48
Leu Ser Pro Ala Gln Ala Ser Gly Ile Trp Cys Arg Asp Ser Ala Phe
                               -40                               -35                               -30
tcc ggg tgg agc gag cgt gga tcc agt tcg cgg cgg ggt tgt ttg ggt      96

```

70

Ser	Gly	Trp	Ser	Glu	Arg	Gly	Ser	Ser	Ser	Arg	Arg	Gly	Cys	Leu	Gly		
		-25					-20					-15					
caa	gtt	gct	gct	gct	gct	gcg	gct	gcg	aat	gct	gct	tct	cct	cct	cct		144
Gln	Val	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Asn	Ala	Ala	Ser	Pro	Pro	Pro		
		-10					-5				1				5		
tca	gcg	aac	agc	ctg	gga	tcg	ggt	ggt	cgc	tgc	aag	ctg	agg	gct	ccg		192
Ser	Ala	Asn	Ser	Leu	Gly	Ser	Gly	Gly	Arg	Cys	Lys	Leu	Arg	Ala	Pro		
				10					15					20			
acg	agc	cgc	ccg	tcc	cag	tcg	cgg	ccc	cga	tcc	ctg	ccc	cag	gcc	cg		240
Thr	Ser	Arg	Pro	Ser	Gln	Ser	Arg	Pro	Arg	Ser	Leu	Pro	Gln	Ala	Arg		
				25				30					35				
ttc	ccc	ccg	tcg	ccg	ctg	ccg	cct	tct	cct	gca	gga	acg	agt	tgc	aca		288
Phe	Pro	Pro	Ser	Pro	Leu	Pro	Pro	Ser	Pro	Ala	Gly	Thr	Ser	Cys	Thr		
		40					45					50					
tgt	tac	tgc	ttc	cta	cta	ata	aat	gct	gac	ctg	atc	aaa	tgg	agc	cca		336
Cys	Tyr	Cys	Phe	Leu	Leu	Ile	Asn	Ala	Asp	Leu	Ile	Lys	Trp	Ser	Pro		
	55					60					65						
aaa	aaa	aaa	aaa	aaa													351
Lys	Lys	Lys	Lys	Lys													

<210> 84
 <211> 438
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 15..278

<220>
 <221> sig_peptide
 <222> 15..122
 <223> Von Heijne matrix
 score 3.79790641648006
 seq ILMRDFSPSGIFG/AF

<220>
 <221> polyA_site
 <222> 423..438

<220>
 <221> polyA_signal
 <222> 396..401

<400>	84																
accaggactc	caaa	atg	gcg	tca	gtt	gta	cca	gtg	aag	gac	aag	aaa	ctt				50
		Met	Ala	Ser	Val	Val	Pro	Val	Lys	Asp	Lys	Lys	Leu				
			-35					-30					-25				
ctg	gag	gtc	aaa	ctg	ggg	gag	ctg	cca	agc	tgg	atc	ttg	atg	cg	gac		98
Leu	Glu	Val	Lys	Leu	Gly	Glu	Leu	Pro	Ser	Trp	Ile	Leu	Met	Arg	Asp		
			-20					-15					-10				
ttc	agt	cct	agt	ggc	att	ttc	gga	gcg	ttt	caa	aga	ggt	tac	tac	cg		146
Phe	Ser	Pro	Ser	Gly	Ile	Phe	Gly	Ala	Phe	Gln	Arg	Gly	Tyr	Tyr	Arg		
			-5				1					5					
tac	tac	aac	aag	tac	atc	aat	gtg	aag	aag	ggg	agc	atc	tcg	ggg	att		194
Tyr	Tyr	Asn	Lys	Tyr	Ile	Asn	Val	Lys	Lys	Gly	Ser	Ile	Ser	Gly	Ile		
		10				15				20							
acc	atg	gtg	ctg	gca	tgc	tac	gtg	ctc	ttt	agc	tac	tcc	ttt	tcc	tac		242
Thr	Met	Val	Leu	Ala	Cys	Tyr	Val	Leu	Phe	Ser	Tyr	Ser	Phe	Ser	Tyr		
	25				30				35					40			
aag	cat	ctc	aag	cac	gag	cgg	ctc	cgc	aaa	tac	cac	tgaagaggac					288
Lys	His	Leu	Lys	His	Glu	Arg	Leu	Arg	Lys	Tyr	His						

	45	50	
acactctgca	ccccccacc	ccacgacctt	ggcccgagcc cctccgtgag gaacacaatc 348
tcaatcggtg	ctgaatcctt	tcatatccta	ataggaatta acctccaaat aaaacatgac 408
tggtacgtgt	gctgaaaaaa	aaaaaaaaaa	438

<210> 85
 <211> 419
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 74..259

<220>
 <221> sig_peptide
 <222> 74..199
 <223> Von Heijne matrix
 score 3.56683458636822
 seq CXXSCCSCCPVGC/AK

<220>
 <221> polyA_site
 <222> 404..419

<220>
 <221> polyA_signal
 <222> 388..393

<400> 85	
cactccgcct tccacgtgca cccactgcct cttcccttct cgcttgggaa ctctagtctc	60
gcctcggggtt gca atg gac ccc aac tgc tcc tgt gcc gct gca ggt gtc	109
Met Asp Pro Asn Cys Ser Cys Ala Ala Gly Val	
-40 -35	
tcc tgc acc tgc gcc agc tcc tgc aag tgc aaa gag tgc aaa tgc acc	157
Ser Cys Thr Cys Ala Ser Ser Cys Lys Cys Lys Glu Cys Lys Cys Thr	
-30 -25 -20 -15	
tcc tgc aar rag agc tgc tgc tcc tgc tgc cct gtg ggc tgt gcc aag	205
Ser Cys Lys Xaa Ser Cys Cys Ser Cys Cys Pro Val Gly Cys Ala Lys	
-10 -5 1	
tgt gcc cag ggc tgc atc tgc aaa ggg gca tcg gag aag tgc agc tgc	253
Cys Ala Gln Gly Cys Ile Cys Lys Gly Ala Ser Glu Lys Cys Ser Cys	
5 10 15	
tgc gcc tgatgtcggg acagccctgc tcccaagtac aaatagagtg acccgtaaaa	309
Cys Ala	
20	
tccaggatatt ttgttttttt gctacaatct tgaccccttt gctacattcc tttttttctg	369
tgaaatatgt gaataataat taaacactta gaccaaaaaa aaaaaaaaaa	419

<210> 86
 <211> 396
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 1..396

<220>
 <221> polyA_site
 <222> 381..396

<400> 86

72

ctc ttc ccg gct cca gct ccg ccg cca gct cca gcc ttt gct ccm cct	48
Leu Phe Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Phe Ala Pro Pro	
1 5 10 15	
ccc aaa gtc ccc tcc ccg gag cgg agc gca cct agg gtc cct ctt ccg	96
Pro Lys Val Pro Ser Pro Glu Arg Ser Ala Pro Arg Val Pro Leu Pro	
20 25 30	
tcc ccc cag ccc agc tac ccr ttc aga cca gca gcc tcg ggg ggc acc	144
Ser Pro Gln Pro Ser Tyr Pro Phe Arg Pro Ala Ala Ser Gly Gly Thr	
35 40 45	
ccc ccg cca gcc tgc ctc cct ccc gct cag ccc tgc cag ggt tcc cca	192
Pro Pro Pro Ala Cys Leu Pro Pro Ala Gln Pro Cys Gln Gly Ser Pro	
50 55 60	
gcc atg aat ctc ttc cga ttc ctg gga gac ctc tcc cac ctc ctc gcc	240
Ala Met Asn Leu Phe Arg Phe Leu Gly Asp Leu Ser His Leu Leu Ala	
65 70 75 80	
atc atc ttg cta ctg ctc aaa atc tgg aag tcc cgc tcg tgc gcc gcc	288
Ile Ile Leu Leu Leu Leu Lys Ile Trp Lys Ser Arg Ser Cys Ala Ala	
85 90 95	
cac ccc cag ctt cct ctc tcc ttc tgt ctg tct gtc tgt ctg tct gtc	336
His Pro Gln Leu Pro Leu Ser Phe Cys Leu Ser Val Cys Leu Ser Val	
100 105 110	
tct ctc tct ctc tct gkc tct ctc tct ctc tct ttc tct gtc tca aaa	384
Ser Leu Ser Leu Ser Xaa Ser Leu Ser Leu Ser Phe Ser Val Ser Lys	
115 120 125	
aaa aaa aaa aaa	396
Lys Lys Lys Lys	
130	

<210> 87

<211> 493

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 3..491

<400> 87

aa ccc cac cgc cta act ctc cca act ctg cag tcc gtg acc ttc agg	47
Pro His Arg Leu Thr Leu Pro Thr Leu Gln Ser Val Thr Phe Arg	
1 5 10 15	
tgc cct tca agg tca gaa aaa ctc ggc aag aat atg gtt tct agt ttt	95
Cys Pro Ser Arg Ser Glu Lys Leu Gly Lys Asn Met Val Ser Ser Phe	
20 25 30	
agg gtt tct gaa cta caa gta tta cta ggc ttt gct gga cgg aat aaa	143
Arg Val Ser Glu Leu Gln Val Leu Leu Gly Phe Ala Gly Arg Asn Lys	
35 40 45	
agt gga cgc aag cat gac ctc ctg atg agg gcg ctg cat tta ttg aag	191
Ser Gly Arg Lys His Asp Leu Leu Met Arg Ala Leu His Leu Leu Lys	
50 55 60	
agc ggc tgc agc cct gcg gtt cag att aaa atc cga gaa ttg tat aga	239
Ser Gly Cys Ser Pro Ala Val Gln Ile Lys Ile Arg Glu Leu Tyr Arg	
65 70 75	
cgc cga tat cca cga act ctt gaa gga ctt tct gat tta tcc aca atc	287
Arg Arg Tyr Pro Arg Thr Leu Glu Gly Leu Ser Asp Leu Ser Thr Ile	
80 85 90 95	
aaa tca tcg gtt ttc agt ttg gat ggt ggc tca tca cct gta gaa cct	335
Lys Ser Ser Val Phe Ser Leu Asp Gly Gly Ser Ser Pro Val Glu Pro	
100 105 110	
gac ttg gcc gtg gct gga atc cac tcg ttg cct tcc act tca gtt aca	383
Asp Leu Ala Val Ala Gly Ile His Ser Leu Pro Ser Thr Ser Val Thr	
115 120 125	
cct cac tca cca tcc tct cct gtt ggt tct gtg ctg ctt caa gat act	431

73

```

Pro His Ser Pro Ser Ser Pro Val Gly Ser Val Leu Leu Gln Asp Thr
      130      135      140
aag ccc aca ttt gag atg cag cag cca tct ccc cca att cct cct gtc      479
Lys Pro Thr Phe Glu Met Gln Gln Pro Ser Pro Pro Ile Pro Pro Val
      145      150      155
cat cct gat gtg ca      493
His Pro Asp Val
160

```

<210> 88
 <211> 500
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 3..500

```

<400> 88
aa ccc cac cgc cta act ctc cca act ctg cag tcc gtg acc ttc agg      47
   Pro His Arg Leu Thr Leu Pro Thr Leu Gln Ser Val Thr Phe Arg
   1           5           10           15
tgc cct tca agg tca gaa aaa ctc ggc aag aat atg gtt tct agt ttt      95
Cys Pro Ser Arg Ser Glu Lys Leu Gly Lys Asn Met Val Ser Ser Phe
           20           25           30
agg gtt tct gaa cta caa gta tta cta ggc ttt gct gga cgg aat aaa      143
Arg Val Ser Glu Leu Gln Val Leu Gly Phe Ala Gly Arg Asn Lys
           35           40           45
agt gga cgc aag cat gac ctc ctg atg agg gcg ctg cat tta ttg aag      191
Ser Gly Arg Lys His Asp Leu Leu Met Arg Ala Leu His Leu Leu Lys
           50           55           60
agc ggc tgc agc cct gcg gtt cag att aaa atc cga gaa ttg tat aga      239
Ser Gly Cys Ser Pro Ala Val Gln Ile Lys Ile Arg Glu Leu Tyr Arg
           65           70           75
cgc cga tat cca cga act ctt gaa gga ctt tct gat tta tcc aca atc      287
Arg Arg Tyr Pro Arg Thr Leu Glu Gly Leu Ser Asp Leu Ser Thr Ile
80           85           90           95
aaa tca tcg gtt ttc agt ttg gat ggt ggc tca tca cct gta gaa cct      335
Lys Ser Ser Val Phe Ser Leu Asp Gly Gly Ser Ser Pro Val Glu Pro
           100          105          110
gac ttg gcc gtg gct gga atc cac tcg ttg cct tcc act tca gtt aca      383
Asp Leu Ala Val Ala Gly Ile His Ser Leu Pro Ser Thr Ser Val Thr
           115          120          125
cct cac tca cca tcc tct cct gtt ggt tct gtg ctg ctt caa gat act      431
Pro His Ser Pro Ser Ser Pro Val Gly Ser Val Leu Leu Gln Asp Thr
           130          135          140
aag ccc aca ttt gag atg cag cag cca tct ccc cca att cct cct gtc      479
Lys Pro Thr Phe Glu Met Gln Gln Pro Ser Pro Pro Ile Pro Pro Val
           145          150          155
cat cct gat gtg cag tta aaa      500
His Pro Asp Val Gln Leu Lys
160           165

```

<210> 89
 <211> 544
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 141..494

<220>

<221> polyA_site

<222> 529..544

<220>

<221> polyA_signal

<222> 508..513

<400> 89

```

agtctcgcga taactgcgca ggcgcggacc aaagcgatct cttctgagga tccggcaaga      60
tggcagaagt atagcagaag aagaagcgga ccttcgcgaa gttcacctac cgcggcgtgg      120
acctcgacca gctgctggac atg tcc tac gag cag ctg atg cag ctg tac agt      173
                Met Ser Tyr Glu Gln Leu Met Gln Leu Tyr Ser
                1             5             10
gcg cgc cag cgg cgg cgg ctg aac cgg ggc ctg cgg cgg aag cag cac      221
Ala Arg Gln Arg Arg Arg Leu Asn Arg Gly Leu Arg Arg Lys Gln His
                15             20             25
tcc ctg ctg aag cgc ctg cgc aag gcc aag aag gag gcg ccg ccc atg      269
Ser Leu Leu Lys Arg Leu Arg Lys Ala Lys Lys Glu Ala Pro Pro Met
                30             35             40
gag aag ccg gaa gtg gtg aag acg cac ctg cgg gac atg atc atc cta      317
Glu Lys Pro Glu Val Val Lys Thr His Leu Arg Asp Met Ile Ile Leu
                45             50             55
ccc gag atg gtg ggc agc atg gtg ggc gtc tac aac ggc aag acc ttc      365
Pro Glu Met Val Gly Ser Met Val Gly Val Tyr Asn Gly Lys Thr Phe
60                65             70             75
aac cag gtg gag atc aag ccc gag atg atc ggc cac tac ctg ggc gag      413
Asn Gln Val Glu Ile Lys Pro Glu Met Ile Gly His Tyr Leu Gly Glu
                80             85             90
ttc tcc atc acc tac aag ccc gta aag cat ggc cgg ccc ggc atc ggg      461
Phe Ser Ile Thr Tyr Lys Pro Val Lys His Gly Arg Pro Gly Ile Gly
                95             100             105
gcc acc cac tcc tcc cgc ttc atc cct ctc aag taatggctca gctaataaag      514
Ala Thr His Ser Ser Arg Phe Ile Pro Leu Lys
                110             115
gcgcacatga ctccaaaaaa aaaaaaaaaa      544

```

<210> 90

<211> 304

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 2..304

<400> 90

```

a cgt acg gca ggg cgg ggt cgg ccc gca gtg gcc agc tgg gag cta cga      49
  Arg Thr Ala Gly Arg Gly Arg Pro Ala Val Ala Ser Trp Glu Leu Arg
  1             5             10             15
gcg cgg ast tgc gca gaa gac ccc cat cag ggt gcg ggg tgc agt tgc      97
Ala Arg Xaa Cys Ala Glu Asp Pro His Gln Gly Ala Gly Cys Ser Cys
                20             25             30
ggc tcc agg gcc atg gcg gag gag cag ggc cgg gaa cgg gac tcg gtt      145
Gly Ser Arg Ala Met Ala Glu Glu Gln Gly Arg Glu Arg Asp Ser Val
                35             40             45
ccc aag ccg tcg gtg ctg ttc ctc cac cca gac ctg ggc gtg ggc ggc      193
Pro Lys Pro Ser Val Leu Phe Leu His Pro Asp Leu Gly Val Gly Gly
                50             55             60
gct gag cgg ctg gtg ctc ttg cct ttt ccc act gag aga agg ctg ctc      241
Ala Glu Arg Leu Val Leu Leu Pro Phe Pro Thr Glu Arg Arg Leu Leu
65                70             75             80
ttt tgt act gcc ccc cgc tca tta aac agc ctc ccc cta aaa aaa aaa      289
Phe Cys Thr Ala Pro Arg Ser Leu Asn Ser Leu Pro Leu Lys Lys Lys

```

85
 90
 95
 304
 aaa awt gcr aaa agc
 Lys Xaa Ala Lys Ser
 100

<210> 91
 <211> 302
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 3..302

<220>
 <221> polyA_site
 <222> 281..302

<220>
 <221> polyA_signal
 <222> 230..235

<400> 91
 ac gta cgg cag ggc ggg gtc ggc ccg cag tgg cca gct ggg agc tac 47
 Val Arg Gln Gly Gly Val Gly Pro Gln Trp Pro Ala Gly Ser Tyr
 1 5 10 15
 gag cgc gga gct tgc gca gaa gac ccc cat cag ggt gcg ggg tgc agt 95
 Glu Arg Gly Ala Cys Ala Glu Asp Pro His Gln Gly Ala Gly Cys Ser
 20 25 30
 tgc ggc tcc agg gcc atg gcg gag gag cag ggc cgg gaa cgg gac tcg 143
 Cys Gly Ser Arg Ala Met Ala Glu Glu Gln Gly Arg Glu Arg Asp Ser
 35 40 45
 gtt ccc aag ccg tcg gtg ctg ttc ctc cac cca gac ctg ggc gtg ggc 191
 Val Pro Lys Pro Ser Val Leu Phe Leu His Pro Asp Leu Gly Val Gly
 50 55 60
 ggc gct gag cgg ctg gtg ctc ttg cct ttt ccc act gag aga agg ctg 239
 Gly Ala Glu Arg Leu Val Leu Leu Pro Phe Pro Thr Glu Arg Arg Leu
 65 70 75
 ctc ttt tgt act gcc ccc cgc tca tta aac agc ctc ccc cta aaa aaa 287
 Leu Phe Cys Thr Ala Pro Arg Ser Leu Asn Ser Leu Pro Leu Lys Lys
 80 85 90 95
 aaa aaa aat gcg aaa 302
 Lys Lys Asn Ala Lys
 100

<210> 92
 <211> 596
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 154..594

<400> 92
 atagtgcgc gcagacagag ggggtggccag tttgttggtt tacctaaaag gtgttttcca 60
 catggagtgg gactgagagg cacaggtggc tgaggggacc cgcctgggac gtgagggcga 120
 ggtcctcggc cccttcattc gtcagctgca gca atg gaa tac gtg ctg gaa gtg 174
 Met Glu Tyr Val Leu Glu Val
 1 5
 aag aac tct ccg cgg cac ctc ctg aag caa ttc aca gtg tgt gac gtt 222
 Lys Asn Ser Pro Arg His Leu Leu Lys Gln Phe Thr Val Cys Asp Val
 10 15 20

76

```

cct ctg tat gac att tgt gac tac aac gtc tcc aga gac cga tgc cag      270
Pro Leu Tyr Asp Ile Cys Asp Tyr Asn Val Ser Arg Asp Arg Cys Gln
   25                               30                               35
gag ctg ggg tgc tgc tty tac gaa ggc gty tgc tac aag aaa gcg gtt      318
Glu Leu Gly Cys Cys Phe Tyr Glu Gly Val Cys Tyr Lys Lys Ala Val
   40                               45                               50                               55
ccc att tac atc cac gtg ttc tct gcc ttg att gtg atc aaa gct ggg      366
Pro Ile Tyr Ile His Val Phe Ser Ala Leu Ile Val Ile Lys Ala Gly
                               60                               65                               70
gcc ttc gtc atc acc atc atc tac aga gtc att cag gag agc agg aaa      414
Ala Phe Val Ile Thr Ile Ile Tyr Arg Val Ile Gln Glu Ser Arg Lys
                               75                               80                               85
gaa aag gsc atc cct gtg tat gtc gcg ctg cca cag aag tcc agc gaa      462
Glu Lys Xaa Ile Pro Val Tyr Val Ala Leu Pro Gln Lys Ser Ser Glu
   90                               95                               100
aag gcg gag ttg gcc tca tcc agc agc aag tta ggg ctg aag cct gcg      510
Lys Ala Glu Leu Ala Ser Ser Ser Ser Lys Leu Gly Leu Lys Pro Ala
   105                               110                               115
agt cct ggg cct cca agt gct ggg ccc tcg atg aag agt gac gag gat      558
Ser Pro Gly Pro Pro Ser Ala Gly Pro Ser Met Lys Ser Asp Glu Asp
   120                               125                               130                               135
aag gat gat gta aca ggg aca ata aca gaa gcc gaa ga      596
Lys Asp Asp Val Thr Gly Thr Ile Thr Glu Ala Glu
                               140                               145

```

<210> 93

<211> 596

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 154..594

<400> 93

```

atagtgagcg gcagacagag ggggtggccag tttgttgttt tacctaaaag gtgttttcca      60
catggagtgg gactgagagg cacaggtggc tgaggggacc cgccctgggac gtgaggcgca      120
ggtcctcggc cccttcatct gtcagctgca gca atg gaa tac gtg ctg gaa gtg      174
                               Met Glu Tyr Val Leu Glu Val
                               1                               5
aag aac tct ccg cgg cac ctc ctg aag caa ttc aca gtg tgt gac gtt      222
Lys Asn Ser Pro Arg His Leu Leu Lys Gln Phe Thr Val Cys Asp Val
   10                               15                               20
cct ctg tat gac att tgt gac tac aac gtc tcc aga gac cga tgc cag      270
Pro Leu Tyr Asp Ile Cys Asp Tyr Asn Val Ser Arg Asp Arg Cys Gln
   25                               30                               35
gag ctg ggg tgc tgc ttc tac gaa ggc gtc tgc tac aag aaa gcg gtt      318
Glu Leu Gly Cys Cys Phe Tyr Glu Gly Val Cys Tyr Lys Lys Ala Val
   40                               45                               50                               55
ccc att tac atc cac gtg ttc tct gcc ttg att gtg atc ata gct ggg      366
Pro Ile Tyr Ile His Val Phe Ser Ala Leu Ile Val Ile Ile Ala Gly
                               60                               65                               70
gcc ttc gtc atc acc atc atc tac aga gtc att cag gag agc agg aaa      414
Ala Phe Val Ile Thr Ile Ile Tyr Arg Val Ile Gln Glu Ser Arg Lys
   75                               80                               85
gaa aag gcc atc cct gtg tat gtc gcg ctg cca cag aag tcc agc gaa      462
Glu Lys Ala Ile Pro Val Tyr Val Ala Leu Pro Gln Lys Ser Ser Glu
   90                               95                               100
aag gcg gag ttg gcc tca tcc agc agc aag tta ggg ctg aag cct gcg      510
Lys Ala Glu Leu Ala Ser Ser Ser Ser Lys Leu Gly Leu Lys Pro Ala
   105                               110                               115
agt cct ggg cct cca agt gct ggg ccc tcg atg aag agt gac gag gat      558
Ser Pro Gly Pro Pro Ser Ala Gly Pro Ser Met Lys Ser Asp Glu Asp

```

77

```

120          125          130          135
aag gat gat gta aca ggg aca ata aca gaa gcc gaa ga
Lys Asp Asp Val Thr Gly Thr Ile Thr Glu Ala Glu
          140          145

```

```
<210> 94
<211> 1413
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 24..470
```

```
<220>
<221> polyA_site
<222> 1398..1413
```

<400> 94																	
gggttcccggt	tccccgcgga	gcc	atg	cgg	tac	aac	gag	aag	gag	ctg	cag	gct					53
			Met	Arg	Tyr	Asn	Glu	Lys	Glu	Leu	Gln	Ala					
			1				5					10					
ctg	tcc	cgg	cag	cgg	gcc	gag	atg	gcg	gcc	gag	ctg	ggc	atg	agg	ggc		101
Leu	Ser	Arg	Gln	Pro	Ala	Glu	Met	Ala	Ala	Glu	Leu	Gly	Met	Arg	Gly		
			15					20					25				
ccc	aag	aag	ggc	agc	gtg	ctg	aag	cgg	cgg	ctg	gtg	aag	ctg	gtg	gtg		149
Pro	Lys	Lys	Gly	Ser	Val	Leu	Lys	Arg	Arg	Leu	Val	Lys	Leu	Val	Val		
			30					35					40				
aat	ttc	ctc	ttc	tac	ttt	cgg	aca	gac	gag	gcc	gag	ccc	gtc	gga	gcc		197
Asn	Phe	Leu	Phe	Tyr	Phe	Arg	Thr	Asp	Glu	Ala	Glu	Pro	Val	Gly	Ala		
			45				50					55					
ctg	ctg	ctg	gag	cgc	tgc	aga	gtc	gtc	cgg	gaa	gag	ccc	ggc	acc	ttc		245
Leu	Leu	Leu	Glu	Arg	Cys	Arg	Val	Val	Arg	Glu	Glu	Pro	Gly	Thr	Phe		
			60			65					70						
tcc	atc	agc	ttc	att	gag	gac	cct	gag	agg	aag	tat	cac	ttt	gag	tgc		293
Ser	Ile	Ser	Phe	Ile	Glu	Asp	Pro	Glu	Arg	Lys	Tyr	His	Phe	Glu	Cys		
			75		80					85			90				
agc	agc	gag	gag	cag	tgt	cag	gag	tgg	atg	gag	gct	ctg	cgt	cgg	gcc		341
Ser	Ser	Glu	Glu	Gln	Cys	Gln	Glu	Trp	Met	Glu	Ala	Leu	Arg	Arg	Ala		
				95				100					105				
agc	tac	gag	ttc	atg	cgg	aga	agc	ctc	atc	ttc	tac	agg	aac	gaa	atc		389
Ser	Tyr	Glu	Phe	Met	Arg	Arg	Ser	Leu	Ile	Phe	Tyr	Arg	Asn	Glu	Ile		
			110					115				120					
cgg	aag	gtg	acg	ggc	aag	gac	ccc	ctg	gaa	cag	ttc	ggc	ata	tcc	gag		437
Arg	Lys	Val	Thr	Gly	Lys	Asp	Pro	Leu	Glu	Gln	Phe	Gly	Ile	Ser	Glu		
			125			130					135						
gag	gcc	agg	ttc	cag	ctg	agt	ggc	ttg	cag	gcg	tgagcgcagg	gcacgggtggt					490
Glu	Ala	Arg	Phe	Gln	Leu	Ser	Gly	Leu	Gln	Ala							
			140			145											
cagcgtgcag	cgggacggga	ctggccctgc	ccagccatga	atcgcttggc	catgcctgga												550
tctgttttgt	tttggttttt	ggtttttggg	tcagggtttc	actgtgttgc	ccaggctaga												610
gtgcagtggt	gccacagctc	actgtgacct	tgaccttctg	gactcaagtg	atcctcctgc												670
ctcagcttcc	caagtagcgt	ggatcacagg	catgagtcgc	cacaccggc	catcacacct												730
ggatttttcag	tgggaggttt	ttggttttga	gacatccaaa	gcttgaagcc	aggtgggtgt												790
gggcaggggc	tgcattttat	gaaactggcc	agcaagctgc	gctccctggg	gccccggat												850
ccacctaaact	ggcttc																

caaaggttgc agtgagctga gatcacgcc a ctgcactcca gcctgggcaa cagacctga 1390
ctccatcaaa aaaaaaaaaa aaa 1413

<210> 95
<211> 519
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 211..519

<220>
<221> polyA_site
<222> 504..519

<400> 95
agattcttaa ctacgtgtct tacagcaaag ttgcccatgg gaacaaaggc tgtaggggtgt 60
cattgatctt gtaatacaac acatatttga tgggtgaatc ttattagcac tcctttgaaa 120
atgttccagt atggagtcag ttgctgataa agaggaagtt cacagctagg aatccgaagt 180
taagatttct caattgccgc ctgggcactc atg cct gta gtc cca gca ctt ggg 234
Met Pro Val Val Pro Ala Leu Gly
1 5
agg cca agg tgg gca gat cac ctg agg tcg gga gtt cgg gac cag cct 282
Arg Pro Arg Trp Ala Asp His Leu Arg Ser Gly Val Arg Asp Gln Pro
10 15 20
ggc caa cct ggt gaa gcc ccc cca tct cta cta aaa ata caa aaa ttg 330
Gly Gln Pro Gly Glu Ala Pro Pro Ser Leu Leu Lys Ile Gln Lys Leu
25 30 35 40
gcc ggg tat ggt ggt ggc tgc ctg tgg tcc cag cta ctc ggg agg ctg 378
Ala Gly Tyr Gly Gly Gly Cys Leu Trp Ser Gln Leu Leu Gly Arg Leu
45 50 55
agg cgg gag aat cac ttg agc ccg gga ggc gga ggt tgc agt gag ccg 426
Arg Arg Glu Asn His Leu Ser Pro Gly Gly Gly Gly Cys Ser Glu Pro
60 65 70
aga ttg tgc cac tgc act cca gcc tgg gtg aca gag caa gac tcc atc 474
Arg Leu Cys His Cys Thr Pro Ala Trp Val Thr Glu Gln Asp Ser Ile
75 80 85
tca aaa ata gaa aag att tct caa ttg cca aaa aaa aaa aaa aaa 519
Ser Lys Ile Glu Lys Ile Ser Gln Leu Pro Lys Lys Lys Lys Lys
90 95 100

<210> 96
<211> 347
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 73..306

<220>
<221> polyA_site
<222> 332..347

<220>
<221> polyA_signal
<222> 305..310

<400> 96
actccgcctt ccacgtgcac ccactgcctc ttcccttctc gcttggaac tctagtctcg 60
cctcgggttg ca atg gac ccc aac tgc tcc tgt gcc gct ggt gtc tcc tgc 111
Met Asp Pro Asn Cys Ser Cys Ala Ala Gly Val Ser Cys

```

      1           5           10
acc tgc gcc agc tcc tgc aag tgc aaa gag tgc aaa tgc acc tcc tgc      159
Thr Cys Ala Ser Ser Cys Lys Cys Lys Glu Cys Lys Cys Thr Ser Cys
      15           20           25
aag aag agc tgc tgc tcc tgg tgc act ggc ccc acc ctc gtg gac acc      207
Lys Lys Ser Cys Cys Ser Trp Cys Thr Gly Pro Thr Leu Val Asp Thr
      30           35           40           45
tgc cct gcc ctg cca cct gtc tgt ctg tcc caa aga agt tct ggt atg      255
Cys Pro Ala Leu Pro Val Cys Leu Ser Gln Arg Ser Ser Gly Met
      50           55           60
aac ttg agg aca cat gtc cag tgg gag gtg aga cca cct ctc aat att      303
Asn Leu Arg Thr His Val Gln Trp Glu Val Arg Pro Pro Leu Asn Ile
      65           70           75
caa taaagctgct gagaatctag cctcgaaaaa aaaaaaaaaa a      347
Gln

```

<210> 97
 <211> 477
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 120..395

<220>
 <221> polyA_site
 <222> 462..477

```

<400> 97
gcttcttttct agtgtggacc agaggtgggg gcaatgagta aacatcttca ctgtaagggt      60
tttgggactg cccctcaggg cttgaggtct cggcatcgtg tctcagtcag tccttggcg      119
atg tcc cct cag acg cac tca cag act tgc ata cgt aat tta gta aca      167
Met Ser Pro Gln Thr His Ser Gln Thr Cys Ile Arg Asn Leu Val Thr
      1           5           10           15
tgc atc aac tat cct aga aca tct aca ggg tgc aaa gga acc acc act      215
Cys Ile Asn Tyr Pro Arg Thr Ser Thr Gly Cys Lys Gly Thr Thr Thr
      20           25           30
cag aga atc atg gag cca gtg gag tta gaa gtg gaa ggg aca gaa caa      263
Gln Arg Ile Met Glu Pro Val Glu Leu Glu Val Glu Gly Thr Glu Gln
      35           40           45
gac aat gct aaa acc tgt ggt tca cta gga agg ggg aat gaa aac acc      311
Asp Asn Ala Lys Thr Cys Gly Ser Leu Gly Arg Gly Asn Glu Asn Thr
      50           55           60
atg ctc cga ggt gga ttc agc atg aac aca act gtg ggg caa gga att      359
Met Leu Arg Gly Gly Phe Ser Met Asn Thr Thr Val Gly Gln Gly Ile
      65           70           75           80
tcc aag caa aca cac cac act agt acc act tct tcc taaacagaaa      405
Ser Lys Gln Thr His His Thr Ser Thr Ser Ser
      85           90
agggaaagag catccttoga sttccccctc cctccatcag caccgggttt tcctccaaaa      465
aaaaaaaaaa aa      477

```

<210> 98
 <211> 363
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 117..269

<220>

<221> polyA_site

<222> 348..363

<220>

<221> polyA_signal

<222> 329..334

<400> 98

```

cctttctcgt tccccggcca tcttatcggc tgctgttggt tggggggccgt cccgctccta      60
aggcaggaag atggtggccg caaagaagac ggaaatctga aatagagtac tatgct atg      119
                                         Met
                                         1
ttg gct aaa act ggt gtc cat cac tac agt ggc aat aat att gaa ctg      167
Leu Ala Lys Thr Gly Val His His Tyr Ser Gly Asn Asn Ile Glu Leu
                    5                      10                      15
ggc aca gca tgc gga aaa tac tac aga gtg tgc aca ctg gct atc att      215
Gly Thr Ala Cys Gly Lys Tyr Tyr Arg Val Cys Thr Leu Ala Ile Ile
                    20                      25                      30
gat cca ggt gac tct gac atc att aga agc atg cca gaa cag act ggt      263
Asp Pro Gly Asp Ser Asp Ile Ile Arg Ser Met Pro Glu Gln Thr Gly
                    35                      40                      45
gaa aag taaacctttt cacctacaaa atttcacctg caaaccttaa acctgcaaaa      319
Glu Lys
50
ttttccttta ataaaatttg cttgttctaa aaaaaaaaaa aaaa      363

```

<210> 99

<211> 606

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 115..411

<220>

<221> polyA_site

<222> 591..606

<220>

<221> polyA_signal

<222> 573..578

<400> 99

```

aagaaagggg tgaggcctaa gggacaatca ggatgttttt cagagagaag tgtggatgct      60
ggacaggaag aaccacagat accagatacg ggtactgttg taactctggt ctcc atg      117
                                         Met
                                         1
aaa aaa aag gaa gaa aca aca ctt tca gag atg gag cct gtt gag cca      165
Lys Lys Lys Glu Glu Thr Thr Leu Ser Glu Met Glu Pro Val Glu Pro
                    5                      10                      15
cag tac caa cta gtc aat gct gaa tcg act tct ccc ttt cta cat tgc      213
Gln Tyr Gln Leu Val Asn Ala Glu Ser Thr Ser Pro Phe Leu His Cys
                    20                      25                      30
ctg aga gaa gtc att ggg gaa tac tct gta cac gaa ttt tca ctg ttg      261
Leu Arg Glu Val Ile Gly Glu Tyr Ser Val His Glu Phe Ser Leu Leu
                    35                      40                      45
ggg aaa aca gag agt caa ggg att gga ttg tgg att gca ttg gtg gtt      309
Gly Lys Thr Glu Ser Gln Gly Ile Gly Leu Trp Ile Ala Leu Val Val
50                      55                      60                      65
ttc ctc agt ttc ctc atc ttc tcc aca agt ttc tac ata tcg aat gca      357
Phe Leu Ser Phe Leu Ile Phe Ser Thr Ser Phe Tyr Ile Ser Asn Ala
                    70                      75                      80

```

gag cag ccc ttc ttc aaa gaa cct cct acg gaa gct gct aag gaa ctc 405
 Glu Gln Pro Phe Phe Lys Glu Pro Pro Thr Glu Ala Ala Lys Glu Leu
 85 90 95
 agt ctg tagctctgcg tggagccatg tgtaaact gaactgagac ctgccacctc 461
 Ser Leu
 ctactaccta agggcccatt ttcattctgat atcatccccc agaaacaaac tcatgatgac 521
 ttccatgttt tttttagatt agatacatgg agaattttcc tttcccttag aattaaaatc 581
 ctgcattcta aaaaaaaaaa aaaat 606

<210> 100
 <211> 648
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 125..508

<220>
 <221> polyA_site
 <222> 633..648

<220>
 <221> polyA_signal
 <222> 610..615

<400> 100
 agctgcctgg gtaaggccca agatggctgt cttgcctta gtactcgtgt gaagttggca 60
 gggacgggttc ctgtcatctt cttgggctta tttggtgtgc tgttgaaggg gggagactag 120
 agaa atg gca ggg aac ctc tta tcc ggg gca ggt agg cgc ctg tgg gac 169
 Met Ala Gly Asn Leu Leu Ser Gly Ala Gly Arg Arg Leu Trp Asp
 1 5 10 15
 tgg gtg cct ctg gcg tgc aga agc ttc tct ctt ggt gtg cct aga ttg 217
 Trp Val Pro Leu Ala Cys Arg Ser Phe Ser Leu Gly Val Pro Arg Leu
 20 25 30
 atc ggt ata agg ctc act ctc ccg ccc ccc aaa gtg gtt gat cgt tgg 265
 Ile Gly Ile Arg Leu Thr Leu Pro Pro Pro Lys Val Val Asp Arg Trp
 35 40 45
 aac gag aaa agg gcc atg ttc gga gtg tat gac aac atc ggg atc ctg 313
 Asn Glu Lys Arg Ala Met Phe Gly Val Tyr Asp Asn Ile Gly Ile Leu
 50 55 60
 gga aac ttt gaa aag cac ccc aaa gaa ctg atc agg ggg ccc ata tgg 361
 Gly Asn Phe Glu Lys His Pro Lys Glu Leu Ile Arg Gly Pro Ile Trp
 65 70 75
 ctt cga ggt tgg aaa ggg aat gaa ttg caa cgt tgt atc cga aag agg 409
 Leu Arg Gly Trp Lys Gly Asn Glu Leu Gln Arg Cys Ile Arg Lys Arg
 80 85 90 95
 aaa atg gtt gga agt aga atg ttc gct gat gac ctg cac aac ctt aat 457
 Lys Met Val Gly Ser Arg Met Phe Ala Asp Asp Leu His Asn Leu Asn
 100 105 110
 aaa cgc atc cgc tat ctc tac aaa cac ttt aac cga cat ggg aag ttt 505
 Lys Arg Ile Arg Tyr Leu Tyr Lys His Phe Asn Arg His Gly Lys Phe
 115 120 125
 cga tagaagagaa agctgagaac ttcggaaaag gctcatctgt caccctggag 558
 Arg
 aagggaact gtacttttcc ctgtgaggaa acggctttgt attttctctg taataaaatg 618
 gggcttcttt ggacaaaaaa aaaaaaaaaa 648

<210> 101
 <211> 491
 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS

<222> 50..205

<220>

<221> polyA_site

<222> 469..491

<400> 101

```

actctttcct tggctgctca ggtcataggt gctgtgatct agagacaaa atg ttc ctt      58
                                   Met Phe Leu
                                   1
aca ctg gca gat act tgc aaa cta agg gga atg agc ttc ctt ctg aat      106
Thr Leu Ala Asp Thr Cys Lys Leu Arg Gly Met Ser Phe Leu Leu Asn
   5                               10                               15
gtt tat gaa gga gag gcc act gtg tca tct gtc tta gag cta ttg gaa      154
Val Tyr Glu Gly Glu Ala Thr Val Ser Ser Val Leu Glu Leu Leu Glu
  20                               25                               30                               35
tcc tgg atc att gtg gga aat gaa aga tac ttc gat gga atc agc agc      202
Ser Trp Ile Ile Val Gly Asn Glu Arg Tyr Phe Asp Gly Ile Ser Ser
              40                               45                               50
cat tgatccaatg ccaactccaa gactggaacg tcgcaatgat agttccaagg      255
His
cggaaatttg acgtaattct tttcgacaca gttttacagg tctggatgcc cattttaatt      315
cttctgaaag catgcctcct ccttctggct tcaggactcc atctccagcc tcttgatcta      375
aaaataatcc ccaaaccaaa aaattagata ctatttcctc aaaattaggt attttaatca      435
aacatcttta acataaatac attattatca cccaaaaaaa aaaaaaaatg cgaaaa      491

```

<210> 102

<211> 565

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 90..341

<220>

<221> polyA_site

<222> 534..565

<220>

<221> polyA_signal

<222> 515..520

<400> 102

```

gttgagggag tcgggcccgcg actgtggtcg tttttataacc ttcccgcgcg gacgccggcg      60
ctgccaacgg aagggcgggc tctgctgcc atg aag gtg aag att aag tgc tgg      113
                                   Met Lys Val Lys Ile Lys Cys Trp
                                   1                               5
aac ggc gtg gcc act tgg ctc tgg gtg gcc aac gat gag aac tgt ggc      161
Asn Gly Val Ala Thr Trp Leu Trp Val Ala Asn Asp Glu Asn Cys Gly
  10                               15                               20
atc tgc agg atg gca ttt aac gga tgc tgc cct gac tgc aag gtg ccc      209
Ile Cys Arg Met Ala Phe Asn Gly Cys Cys Pro Asp Cys Lys Val Pro
  25                               30                               35                               40
ggc gac gac tgc ccg ctg gtg tgg ggc cag tgc tcc cac tgc ttc cac      257
Gly Asp Asp Cys Pro Leu Val Trp Gly Gln Cys Ser His Cys Phe His
              45                               50                               55
atg cat tgc atc ctc aag tgg ctg cac gca cag cag gtg cag cag cac      305
Met His Cys Ile Leu Lys Trp Leu His Ala Gln Gln Val Gln Gln His
              60                               65                               70
tgc ccc atg tgc cgc cag gaa tgg aag ttc aag gag tgaggcccga      351

```

Cys Pro Met Cys Arg Gln Glu Trp Lys Phe Lys Glu

75

80

```
cctggctctc gctggagggg catcctgaga ctcttctctc atgctggcgc cgatggctgc 411
tggggacagc gccctgagc tgcaacaagg tggaaacaag ggctggagct gcgtttgttt 471
tgccatcact atgttgacac ttttatccaa taagtgtaaa ctcatataac tactcaaadc 531
ttaaaaaaaaa aaaaaaaagaa aaaaaaaaaa aaaa 565
```

<210> 103

<211> 661

<212> DNA

<213> Homo sapiens

 $\langle 220 \rangle$

<221> CDS

<222> 169..420

<220>

<221> polyA_site

 $\langle 222 \rangle \quad 614 \dots \overline{661}$ $\langle 220 \rangle$

<221> polyA_signal

<222> 594..599

<400> 103

```

gagactcggc  gggcgctgtt  gagggagtcg  gcccgcgact  gtggtcgttt  ttataccttc      60
ccgcgcggac  gccggcctgc  caacggaagg  gcggagacgg  agtttcgtca  tgttggccag     120
gcccat ttgga  gatctttgaa  gatatcctca  acgtgaggct  ctgctgcc  atg  aag  gtg     177
                               Met  Lys  Val
                               1

```

aag att aag tgc tgg aac ggc gtg gcc act tgg ctc tgg gtg gcc aac 225
Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp Val Ala Asn
5 10 15

gat gag aac tgt ggc atc tgc agg atg gca ttt aac gga tgc tgc cct 273
Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly Cys Cys Pro
20 25 30 35

gac tgc aag gtg ccc ggc gac gac tgc ccg ctg gtg tgg ggc cag tgc 321
Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp Gly Gln Cys
40 45 50

tcc cac tgc ttc cac atg cat tgc atc ctc aag tgg ctg cac gca cag 369
Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu His Ala Gln
55 60 65

cag gtg cag cag cac tgc ccc atg tgc cgc cag gaa tgg aag ttc aag 417
Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp Lys Phe Lys
70 75 80

gag tgaggcccgga cctggctctc gctggagggg catcctgaga ctcttctctc	470
Glu	
atgctggcgc cgatggctgc tggggacagc gccctgagc tgcaacaagg tggaaacaag	530
ggctggagct gcgtttgttt tgccatcact atgttgacac ttttatccaa taagtgaaaa	590
ctcattaaac tactcaaatc ttgaaaaaaaa aaaaaaaaaa aaaagaaaaa	650
aaaaaaaaaa a	661

$\langle 210 \rangle$ 104

<211> 609

<212> DNA

<213> Homo sapiens

 $\langle 220 \rangle$

<221> CDS

<222> 144..395

<220>

<221> polyA site

<222> 594..609

<220>

<221> polyA_signal

<222> 569..574

<400> 104

```

gtttttatac cttcccgccg ggaacgccggc gctgcccaacg gaagggcggg tagggcggtg      60
cgtgattaga gacggagttt cgtcatgttg gccaggccca ttgagatct ttgaagatat      120
cctcaacgtg aggctctgct gcc atg aag gtg aag att aag tgc tgg aac ggc      173
                Met Lys Val Lys Ile Lys Cys Trp Asn Gly
                1             5             10
gtg gcc act tgg ctc tgg gtg gcc aac gat gag aac tgt ggc atc tgc      221
Val Ala Thr Trp Leu Trp Val Ala Asn Asp Glu Asn Cys Gly Ile Cys
                15             20             25
agg atg gca ttt aac gga tgc tgc cct gac tgc aag gtg ccc ggc gac      269
Arg Met Ala Phe Asn Gly Cys Cys Pro Asp Cys Lys Val Pro Gly Asp
                30             35             40
gac tgc ccg ctg gtg tgg ggc cag tgc tcc cac tgc ttc cac atg cat      317
Asp Cys Pro Leu Val Trp Gly Gln Cys Ser His Cys Phe His Met His
                45             50             55
tgc atc ctc aag tgg ctg cac gca cag cag gtg cag cag cac tgc ccc      365
Cys Ile Leu Lys Trp Leu His Ala Gln Gln Val Gln Gln His Cys Pro
                60             65             70
atg tgc cgc cag gaa tgg aag ttc aag gag tgaggcccg cctggctctc      415
Met Cys Arg Gln Glu Trp Lys Phe Lys Glu
                75             80
gctggagggg catcctgaga ctcccttcctc atgctggcgc cgatggctgc tggggacagc      475
gcccttgagc tgcaacaagg tggaaacaag ggctggagct gcgtttgttt tgccatcact      535
atgttgacac ttttatccaa taagtgaataa ctcatataac tactcaaatac ttgaaaagaa      595
aaaaaaaaaa aaaa      609

```

<210> 105

<211> 635

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 177..428

<220>

<221> polyA_site

<222> 620..635

<220>

<221> polyA_signal

<222> 602..607

<400> 105

```

aagactcggc gggcgctggt gagggagtcg ggccgcgact gtggctcgtt ttataccttc      60
ccgcgcggac gccggcgctg ccaacggaag ggcgggtagg gcgagacgga gtttcgcat      120
gttggccagg cccatttgag atctttgata tatcctcaac gtgaggctct gctgcc atg      179
                Met
                1
aag gtg aag att aag tgc tgg aac ggc gtg gcc act tgg ctc tgg gtg      227
Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp Val
                5             10             15
gcc aac gat gag aac tgt ggc atc tgc agg atg gca ttt aac gga tgc      275
Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly Cys
                20             25             30
tgc cct gac tgc aag gtg ccc ggc gac gac tgc ccg ctg gtg tgg ggc      323
Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp Gly

```

```

      35              40              45
cag tgc tcc cac tgc ttc cac atg cat tgc atc ctc aag tgg ctg cac      371
Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu His
50              55              60              65
gca cag cag gtg cag cag cac tgc ccc atg tgc cgc cag gaa tgg aag      419
Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp Lys
      70              75              80
ttc aag gag tgaggcccgga cctggctctc gctggagggg catcctgaga      468
Phe Lys Glu
ctccttcctc atgctggcgc cgatggctgc tggggacagc gccctgagc tgcaacaagg      528
tggaacaag ggctggagct gcgtttgttt tgccatcact atgttgacac ttttatccaa      588
taagtgaaaa ctcatataac tactcaaatc caaaaaaaaa aaaaaaa      635

```

<210> 106
 <211> 573
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 114..365

<220>
 <221> polyA_site
 <222> 558..573

<220>
 <221> polyA_signal
 <222> 539..544

```

<400> 106
ggcgtgcgag actcggcggg cgctgttgag ggagtcgggc cgcgactgtg gtcgttttta      60
taccttcccg cgcggacgcc ggcgctgcc aacgaagggc gggctctgct gcc atg      116
                                     Met
                                     1
aag gtg aag att aag tgc tgg aac ggc gtg gcc act tgg ctc tgg gtg      164
Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp Val
      5              10              15
gcc aac gat gag aac tgt ggc atc tgc agg atg gca ttt aac gga tgc      212
Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly Cys
      20              25              30
tgc cct gac tgc aag gtg ccc ggc gac gac tgc ccg ctg gtg tgg ggc      260
Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp Gly
      35              40              45
cag tgc tcc cac tgc ttc cac atg cat tgc atc ctc aag tgg ctg cac      308
Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu His
50              55              60              65
gca cag cag gtg cag cag cac tgc ccc atg tgc cgc cag gaa tgg aag      356
Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp Lys
      70              75              80
ttc aag gag tgaggcccgga cctggctctc gctggagggg catcctgaga      405
Phe Lys Glu
ctccttcctc atgctggcgc cgatggctgc tggggacagc gccctgagc tgcaacaagg      465
tggaacaag ggctggagct gcgtttgttt tgccatcact atgttgacac ttttatccaa      525
taagtgaaaa ctcatataac tactcaaatc ctaaaaaaaaa aaaaaaa      573

```

<210> 107
 <211> 689
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS

<222> 44..283

<220>

<221> polyA_site

<222> 674..689

<220>

<221> polyA_signal

<222> 644..649

<400> 107

```

tttcgccgcc aaagcatcca gcagccccct gctccggccc agc atg gcg acc ccg      55
                                     Met Ala Thr Pro
                                     1
acc cag acc ccc aca aag gct cct gag gaa cct gac cca ttt tac tat      103
Thr Gln Thr Pro Thr Lys Ala Pro Glu Glu Pro Asp Pro Phe Tyr Tyr
5                                     10                                     15                                     20
gac tac aac acg gtg cag act gtg ggc atg act ctg gca acc atc ttg      151
Asp Tyr Asn Thr Val Gln Thr Val Gly Met Thr Leu Ala Thr Ile Leu
25                                     30                                     35
ttc ctg ctg ggt atc ctc atc gtc atc agc aag aag gtg aag tgc agg      199
Phe Leu Leu Gly Ile Leu Ile Val Ile Ser Lys Lys Val Lys Cys Arg
40                                     45                                     50
aag gcg gac tcc agg tct gag agc cca acc tgc aaa tcc tgt aag tct      247
Lys Ala Asp Ser Arg Ser Glu Ser Pro Thr Cys Lys Ser Cys Lys Ser
55                                     60                                     65
gag ctt ccc tct tca gcc cct ggt ggc ggc ggc gtg taacaccttc      293
Glu Leu Pro Ser Ser Ala Pro Gly Gly Gly Gly Val
70                                     75                                     80
ccgaggaaac tccgctgccg accctgcctg agcgcgggag cctgaggacc ggggtggaggc      353
gggtggggacc cagccgcgcg ccgggagcgc tccccggaat gagccgcccc acccacccca      413
aggctggagc cgctgcaccc tgccgtccct ctccaggcct tggcaatgac gatcccccaa      473
agagcccgtc tgcaccccag acccagggcc tcaggcctcc agctcctggg atccgggagt      533
ccatcccggc ccagcacccc cagcatcccc gtgtatggcc cccctgcacc tccttgtctc      593
atccccgaag atccgtcccc ctggcccctc agtgtccatg tcttgagctt aataaatgtg      653
catttggttt tttcctctgc aaaaaaaaaa aaaaaa      689

```

<210> 108

<211> 561

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 99..479

<220>

<221> polyA_site

<222> 546..561

<220>

<221> polyA_signal

<222> 526..531

<400> 108

```

tttcgaaggg aaggagagccc cctataaaac agcctacagt ggacagtctg gtcggcagag      60
ccgcaggtca gtcgtgaaga gggagctcta ttgccacc atg agt ttc tcc ggc aag      116
                                     Met Ser Phe Ser Gly Lys
                                     1                                     5
tac caa ctg cag agc cag gaa aac ttt gaa gcc ttc atg aag gca atc      164
Tyr Gln Leu Gln Ser Gln Glu Asn Phe Glu Ala Phe Met Lys Ala Ile
10                                     15                                     20
ggc ctg ccg gaa gag ctc atc cag aag ggg aag gat atc aag ggg gtg      212

```

[illegible]

```
<210> 109
<211> 738
<212> DNA
<213> Homo sapiens
```

<220>
<221> CDS
<222> 1..261

```
<220>
<221> polyA_site
<222> 724..738
```

```
<220>  
<221> polyA_signal  
<222> 712..717
```

<400> 109																
gac	gca	gta	gcc	gtg	ggg	aag	ttt	tcg	caa	tgg	ccg	ccg	gaa	cgg	tcg	48
Asp	Ala	Val	Ala	Val	Gly	Lys	Phe	Ser	Gln	Trp	Pro	Pro	Glu	Arg	Ser	
1				5				10				15				
ccg	gcc	gat	gct	ctc	tgc	gac	gga	aag	aaa	gcc	gcc	gga	ccc	act	tcc	96
Pro	Ala	Asp	Ala	Leu	Cys	Asp	Gly	Lys	Lys	Ala	Ala	Gly	Pro	Thr	Ser	
20				25				30								
ggt	gcg	ttt	tac	aat	gaa	ctc	cgg	gca	ctc	gag	tcg	ctc	ccc	gga	aac	144
Gly	Ala	Phe	Tyr	Asn	Glu	Leu	Arg	Ala	Leu	Glu	Ser	Leu	Pro	Gly	Asn	
35				40				45								
tat	agt	tct	gct	tcc	acc	gtt	gtc	cta	ctt	tgg	ata	gca	tca	aga	caa	192
Tyr	Ser	Ser	Ala	Ser	Thr	Val	Val	Leu	Leu	Trp	Ile	Ala	Ser	Arg	Gln	
50				55				60								
aaa	tct	cag	tct	tcg	agg	aac	tca	cag	tgt	agc	aga	aaa	gac	aaa	gga	240
Lys	Ser	Gln	Ser	Ser	Arg	Asn	Ser	Gln	Cys	Ser	Arg	Lys	Asp	Lys	Gly	
65				70				75				80				
ggt	aaa	cag	gag	gag	cat	gaa	taattcatct	tgatgtaatc	atgcaagact							291
Gly	Lys	Gln	Glu	Glu	His	Glu										
85																
cctaagaaga	tgttaggaga	ccctgtcacc	attottttgg	cctctagtgt	cctttatctt											351
catgttttgc	ggcttccttt	tactactatg	gaaatacaac	atatgcaaaa	gtagtgtttt											411
cgctgttttaa	aaacatcgtc	agcagcatca	tcatcattaa	ctgcaccagg	aagccatgac											471
tcctttctgg	acaaggtgca	cttgcctgtc	atcctttccat	ctctttcttc	ctccttgaag											531

ttcacaaat	ccagtcagg	agcattgcag	agaatgctag	gtgtgtctgc	agacttgact	591
ttaaaagaaa	acaagttcct	tcaagtgagt	tcatcatgat	gggaatgata	atttgtttga	651
gagaatgtgt	gttctttggc	aaaggaactc	ctgttataacc	atgacagtaa	taatagcagt	711
aataaaatca	ccaaaaaaaa	aaaaaaaa				738

```
<210> 110
<211> 584
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 88..375
```

```
<220>
<221> polyA_site
<222> 567..584
```

```
<220>  
<221> polyA_signal  
<222> 546..551
```

<400>	110
gagtcagtct cgccagctgc cgggtctttcgg ggggctccgt aacttttctat ccgtccgcgt	60
cagcgcyttg ccaccctcat ctccaat atg cct ggt ccg acc ccc agt ggc act	114
	Met Pro Gly Pro Thr Pro Ser Gly Thr
	1 5
aac gtg gga tcc tca ggg cgc tct ccc agc aaa gca gtg gcc gcc cgg	162
Asn Val Gly Ser Ser Gly Arg Ser Pro Ser Lys Ala Val Ala Ala Arg	
10 15 20 25	
gcg gcg gga tcc act gtc cgg cag agg aaa aat gcc agc tgt ggg aca	210
Ala Ala Gly Ser Thr Val Arg Gln Arg Lys Asn Ala Ser Cys Gly Thr	
30 35 40	
agg agt gca ggc cgc aca acc tcg gca ggc acc ggg ggg atg tgg cga	258
Arg Ser Ala Gly Arg Thr Thr Ser Ala Gly Thr Gly Gly Met Trp Arg	
45 50 55	
ttc tac aca gaa gat tca cct ggg ctc aaa gtt ggc cct gtt cca gta	306
Phe Tyr Thr Glu Asp Ser Pro Gly Leu Lys Val Gly Pro Val Pro Val	
60 65 70	
ttg gtt atg agt ctt ctg ttc atc gct tct gta ttt atg ttg cac att	354
Leu Val Met Ser Leu Leu Phe Ile Ala Ser Val Phe Met Leu His Ile	
75 80 85	
tgg ggc aag tac act cgt tcg tagattcagt tacatccatc tgatcatctga	405
Trp Gly Lys Tyr Thr Arg Ser	
90 95	
agaaggagga aaaaacccaa cattttcttggtg accaaaagta tagtgactat ctgttcatga	465
gagaaatttt ctgtaaagctt gctgtttttac aggggatttta tcaataattg attttgagga	525
atcaqttttt ttctatggct aataaaccttt ttaattcact tataaaaaaa aaaaaaaaaa	584

```
<210> 111
<211> 588
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 16..183
```

```
<220>  
<221> polyA_site  
<222> 289..588
```

 $\langle 220 \rangle$

<221> polyA_signal

<222> 267..272

<400> 111

```

tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg      51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                1         5         10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac      99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                15         20         25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
                30         35         40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
45         50         55
aatattttcc tcctctgatg ctccctcttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcgagc cactgtggat      433
cttttaggac ttgtttggag aaagattgag aaggaaaaga gtgaaaggca gttagtgtgaa      493
gtggaagagt tgaccttcac gtactgcttt ccaaaaaatt gatcctagtt gttattgaat      553
aatgtgtgt ctgttttaag aaaaaaaaaa aaaaaa      588

```

<210> 112

<211> 589

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 16..183

<220>

<221> polyA_site

<222> 289..589

<220>

<221> polyA_signal

<222> 267..272

<400> 112

```

tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg      51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                1         5         10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac      99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                15         20         25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
                30         35         40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
45         50         55
aatattttcc tcctctgatg ctccctcttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcgagc ccactgtgga      433
tcttttagga cttgtttgga gaaagattga gaaggaaaag agtgaaaggc agttagtgtga      493
agtgaagag ttgaccttca tgtactgctt tccaaaaaat tgatcctagt tgttattgaa      553
taaattgttg tctgttttaa gaaaaaaaaa aaaaaa      589

```

<210> 113

<211> 588
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 289..588

<220>
 <221> polyA_signal
 <222> 267..272

<400> 113
 tttattttg cttct atg tca tct tgg agc tta atg tgc atg caa gat atg 51
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
 1 5 10
 gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac 99
 Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
 15 20 25
 ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat 147
 Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
 30 35 40
 aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac 193
 Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
 45 50 55
 aaatattttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc 253
 tgctgttgac aaaaataaat tggagaaaga aaaggaaaaa aaaaaggaag agaaaaagag 313
 agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga 373
 tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggasc cactgtggat 433
 cttttaggac ttgtttggag aaagattgag aaggaaaaaga gtgaaaggca gttagttaga 493
 gtggaagagt tgaccttcat gtactgcttt ccaaaaaatt gatcctagtt gttattgaat 553
 aaattgttgt ctgttttaag aaaaaaaaaa aaaaa 588

<210> 114
 <211> 590
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 289..590

<220>
 <221> polyA_signal
 <222> 267..272

<400> 114
 tttattttg cttct atg tca tct tgg agc tta atg tgc atg caa gat atg 51
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
 1 5 10
 gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac 99
 Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
 15 20 25
 ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat 147
 Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp

30	35	40	
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac			193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro			
45	50	55	
aaatattttcc tcctctgatg ctctctttca gccttttggtg tcctctcctt ctctgcaagc			253
tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag			313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga			373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggagc ccactgtgga			433
tcttttagga cttgttttga gaaagattga gaaggaaaag agtgaaaggc agttagttga			493
agtgaagag ttgaccttca tgtactgctt tccaaaaaat tgatcctagt tgttattgaa			553
taaattgttg tctgttttaa gaaaaaaaaa aaaaaaat			590

<210> 115
 <211> 595
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 572..595

<220>
 <221> polyA_signal
 <222> 552..557

<400> 115	
tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg	51
Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met	
1 5 10	
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac	99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn	
15 20 25	
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat	147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp	
30 35 40	
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac	193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro	
45 50 55	
aaatattttcc tcctctgatg ctctctttca gccttttggtg tcctctcctt ctctgcaagc	253
tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag	313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga	373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggagc ccactgtgga	433
tcttttagga cttgttttga gaaagattga gaaggaaaag agtgaaaggc agttagttga	493
agtgaagag ttgaccttca tgtactgctt tccaaaaaat tgatcctagt tgttattgaa	553
taaattgttg tctgttttaa gaaaaaaaaa aaaaaatgcg aa	595

<210> 116
 <211> 588
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 289..588

<220>

<221> polyA_signal

<222> 267..272

<400> 116

```

tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg      51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                1      5      10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac      99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                15      20      25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
                30      35      40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
                45      50      55
aaatattttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcgagc cactgtggat      433
cttttaggac ttgtttggag aaagattgag aaggaaaaga gtgaaaggca gttagttgaa      493
gtggaagagt tgaccttcac gtactgcttt ccaaaaaatt gatcctagtt gttattgaat      553
aaattgttgt ctgttttaag aaaaaaaaaa aaaaaa      588

```

<210> 117

<211> 588

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 16..183

<220>

<221> polyA_site

<222> 289..588

<220>

<221> polyA_signal

<222> 267..272

<400> 117

```

tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg      51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                1      5      10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac      99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                15      20      25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
                30      35      40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
                45      50      55
aaatattttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcgagc cactgtggat      433
cttttaggac ttgtttggag aaagattgag aaggaaaaga gtgaaaggca gttagttgaa      493
gtggaagagt tgaccttcac gtactgcttt ccaaaaaatt gatcctagtt gttattgaat      553
aaattgttgt ctgttttaag aaaaaaaaaa aaaaaa      588

```

<210> 118
 <211> 588
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 289..588

<220>
 <221> polyA_signal
 <222> 267..272

<400> 118
 tttatatttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg 51
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
 1 5 10
 gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac 99
 Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
 15 20 25
 ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat 147
 Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
 30 35 40
 aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac 193
 Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
 45 50 55
 aaatatttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc 253
 tgctgttgac aaaaataaat tggagaaaga aaaggaaaaa aaaaaggaag agaaaaagag 313
 agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga 373
 tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcgagag cactgtggat 433
 cttttaggac ttgtttggag aaagattgag aaggaaaaga gtgaaaggca gttagttgaa 493
 gtggaagagt tgaccttcac gtactgcttt ccaaaaaatt gatcctagtt gttattgaat 553
 aaattgttgt ctgttttaag aaaaaaaaaa aaaaa 588

<210> 119
 <211> 595
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 571..595

<220>
 <221> polyA_signal
 <222> 551..556

<400> 119
 tttatatttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg 51
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
 1 5 10
 gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac 99
 Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
 15 20 25
 ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat 147

Phe	Arg	Arg	Pro	Gln	Thr	Asp	Gln	Ala	Val	Glu	Phe	Phe	Ile	Arg	Asp		
30						35					40						
aaa	tat	gaa	aag	aag	aaa	tac	tac	gaa	aaa	atg	cca	tagctattac				193	
Lys	Tyr	Glu	Lys	Lys	Lys	Tyr	Tyr	Glu	Lys	Met	Pro						
45					50					55							
aaatattttcc	tcctctgatg	ctcctctttca	gccttttggtgta	tcctctcctt	ctctgcaagc								253				
tgctgttgac	aaaaataaat	tgagagaaaga	aaaggaaaaa	aaaaaggaag	agaaaaagag								313				
agaaaaggag	ccagaaaaagc	cggcaaaaacc	acttacagct	gaaaagctgc	agaagaaaga								373				
tcagcaactg	gagcctaaaa	aaagtaccag	ccctaaaaaa	gctgcgaggc	cactgtggat								433				
cttttaggac	ttgtttgag	aaagattgag	aaggaaaaaga	gtgaaaggca	gttagttgaa								493				
gtggaagagt	tgaccttcac	gtactgcttt	ccaaaaaatt	gatcctagtt	gttattgaat								553				
aaattgttgt	ctgttttaag	aaaaaaaaaa	aaaaatgcga	aa									595				

<210> 120

<211> 601

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 16..183

<220>

<221> polyA_site

<222> 572..601

<220>

<221> polyA_signal

<222> 552..557

<400> 120

tttatttttgc	cttct	atg	tca	tct	tgg	agc	tta	atg	tgc	atg	caa	gat	atg			51	
		Met	Ser	Ser	Trp	Ser	Leu	Met	Cys	Met	Gln	Asp	Met				
		1			5				10								
gga	aat	act	aaa	gca	aga	cta	ctc	tat	gaa	gcc	aat	ctt	cca	gag	aac	99	
Gly	Asn	Thr	Lys	Ala	Arg	Leu	Leu	Tyr	Glu	Ala	Asn	Leu	Pro	Glu	Asn		
		15				20					25						
ttt	cga	aga	cca	cag	aca	gat	caa	gca	gtg	gaa	ttt	ttc	atc	aga	gat	147	
Phe	Arg	Arg	Pro	Gln	Thr	Asp	Gln	Ala	Val	Glu	Phe	Phe	Ile	Arg	Asp		
30					35					40							
aaa	tat	gaa	aag	aag	aaa	tac	tac	gaa	aaa	atg	cca	tagctattac				193	
Lys	Tyr	Glu	Lys	Lys	Lys	Tyr	Tyr	Glu	Lys	Met	Pro						
45					50					55							
aaatattttcc	tcctctgatg	ctcctctttca	gccttttggtgta	tcctctcctt	ctctgcaagc								253				
tgctgttgac	aaaaataaat	tgagagaaaga	aaaggaaaaa	aaaaaggaag	agaaaaagag								313				
agaaaaggag	ccagaaaaagc	cggcaaaaacc	acttacagct	gaaaagctgc	agaagaaaga								373				
tcagcaactg	gagcctaaaa	aaagtaccag	ccctaaaaaa	gctgcgaggc	ccactgtgga								433				
tcttttagga	cttgtttgga	gaaagattga	gaaggaaaag	agtgaaaggc	agttagttga								493				
agtgaagag	ttgaccttca	tgtactgctt	tccaaaaaat	tgatcctagt	tgttattgaa								553				
taaattgttg	tctgttttaa	gaaaaaaaaa	aaaaaatgcg	aaaagctt									601				

<210> 121

<211> 589

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 16..183

<220>

<221> polyA_site

<222> 289..589

<220>

<221> polyA_signal

<222> 267..272

<400> 121

```

tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg      51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                1      5      10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac      99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                15      20      25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
                30      35      40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
45      50      55
aaatattttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggagc ccactgtgga      433
tctttttgga cttgtttgga gaaagattga gaaggaaaag agtgaaaggc agttagttga      493
agtggaagag ttgaccttca tgtactgctt tccaaaaaat tgatcctagt tgttattgaa      553
taaattgttg tctgttttaa gaaaaaaaaa aaaaaa      589

```

<210> 122

<211> 595

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 16..183

<220>

<221> polyA_site

<222> 571..595

<220>

<221> polyA_signal

<222> 551..556

<400> 122

```

tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg      51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                1      5      10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac      99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                15      20      25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
                30      35      40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
45      50      55
aaatattttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggasc cactgtggat      433
cttttaggac ttgtttggag aaagattgag aaggaaaaga gtgaaaggca gttagttgaa      493
gtggaagagt tgaccttcat gtactgcttt ccaaaaaatt gatcctagtt gttattgaat      553
aaattgttgt ctgttttaag aaaaaaaaaa aaaaatgcga aa      595

```


<210> 123
 <211> 588
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 289..588

<220>
 <221> polyA_signal
 <222> 267..272

<400> 123
 tttattttgcttct atg tca tct tgg agc tta atg tgc atg caa gat atg 51
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
 1 5 10
 gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac 99
 Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
 15 20 25
 ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat 147
 Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
 30 35 40
 aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac 193
 Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
 45 50 55
 aaatattttcc tcctctgatg ctctctttca gccttttggtta tcctctcctt ctctgcaagc 253
 tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag 313
 agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga 373
 tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggagc cactgtggat 433
 cttttaggac ttgtttggag aaagattgag aaggaaaaga gtgaaaggca gttagttgaa 493
 gtggaagagt tgaccttcac gtactgcttt ccaaaaaatt gatcctagtt gttattgaat 553
 aaattgttgt ctgttttaag aaaaaaaaaa aaaaaa 588

<210> 124
 <211> 588
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 289..588

<220>
 <221> polyA_signal
 <222> 267..272

<400> 124
 tttattttgcttct atg tca tct tgg agc tta atg tgc atg caa gat atg 51
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
 1 5 10
 gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac 99
 Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
 15 20 25

```

ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
   30                               35                               40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
   45                               50                               55
aaatattttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggagc cactgtggat      433
cttttaggac ttgtttggag aaagattgag aaggaaaaa gtgaaaggca gttagtgtgaa      493
gtggaagagt tgaccttcac gtactgcttt ccaaaaaatt gatcctagtt gttattgaat      553
aaattgttgt ctgttttaag aaaaaaaaaa aaaaaa

```

<210> 125
 <211> 589
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

<220>
 <221> polyA_site
 <222> 289..589

<220>
 <221> polyA_signal
 <222> 267..272

```

<400> 125
tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg      51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                   1               5               10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac      99
Gly Asn Thr Lys Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                   15               20               25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat      147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
   30                               35                               40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac      193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
   45                               50                               55
aaatattttcc tcctctgatg ctctctttca gccttttgta tcctctcctt ctctgcaagc      253
tgctgttgac aaaaataaat tggagaaaga aaaggaaaaa aaaaaggaag agaaaaagag      313
agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga      373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggagc ccactgtgga      433
tcttttagga cttgttttga gaaagattga gaaggaaaag agtgaaaggc agttagtgtga      493
agtggaagag ttgaccttca tgtactgctt tccaaaaaat tgatcctagtt tgttattgaa      553
taaattgttg tctgttttaa gaaaaaaaaa aaaaaa

```

<210> 126
 <211> 663
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 86..451

<220>
 <221> polyA_site

<222> 622..663

$\langle 220 \rangle$

<221> polyA_signal

<222> 597..602

<400> 126

acacatccaa	gcttaagacg	gtgaggtcag	cttcacattc	tcaggaactc	tccttctttg	60
ggccacggaa	ttaacccgag	caggc	atg gag gcc tct	gct ctc acc tca tca	112	
		Met	Glu Ala Ser	Ala Leu Thr Ser Ser		
		1		5		
gca gtg acc agt gtg gcc aaa gtg gtc agg gtg gcc tct ggc tct gcc					160	
Ala Val Thr Ser Val Ala Lys Val Val Arg Val Ala Ser Gly Ser Ala						
10	15		20		25	
gta gtt ttg ccc ctg gcc agg att gct aca gtt gtg att gga gga gtt					208	
Val Val Leu Pro Leu Ala Arg Ile Ala Thr Val Val Ile Gly Gly Val						
	30		35		40	
gtg gcc atg ggc gct gtg ccc atg gtg ctc agt gcc atg ggc ttc act					256	
Val Ala Met Ala Ala Val Pro Met Val Leu Ser Ala Met Gly Phe Thr						
	45		50		55	
gcg gcg gga atc gcc tcg tcc tcc ata gca gcc aag atg atg tcc gcg					304	
Ala Ala Gly Ile Ala Ser Ser Ser Ile Ala Ala Lys Met Met Ser Ala						
	60		65		70	
gcg gcc att gcc aat ggg ggt gga gtt gcc tcg ggc agc ctt gtg gct					352	
Ala Ala Ile Ala Asn Gly Gly Gly Val Ala Ser Gly Ser Leu Val Ala						
	75		80		85	
act ctg cag tca ctg gga gca act gga ctc tcc gga ttg acc aag ttc					400	
Thr Leu Gln Ser Leu Gly Ala Thr Gly Leu Ser Gly Leu Thr Lys Phe						
	90		95		100	
atc ctg ggc tcc att ggg tct gmc att gcg gct gtc att gcg agg ttc					448	
Ile Leu Gly Ser Ile Gly Ser Xaa Ile Ala Ala Val Ile Ala Arg Phe						
	110		115		120	
tac tagctccctg cccctcgccc tgcagagaag agaaccatgc caggggagaa					501	
Tyr						
ggcaccacgc catcctgacc cagcgaggag ccaactatcc caaatataacc tgggggtgaaa					561	
tataccaaat tctgcatctc cagaggaaaa taagaaataa agatgaattg ttgcaactct					621	
aaaaaaaaaa aaaaaaaaaa aaaaacaaaa aaaaaaaaaa aa					663	

<210> 127

<211> 633

<212> DNA

<213> Homo sapiens

 $\langle 220 \rangle$

<221> CDS

<222> 88..453

 $\langle 220 \rangle$

<221> polyA site

$$\langle 222 \rangle \quad \overline{618} \dots \overline{633}$$
 $\langle 220 \rangle$

<221> polyA signal

$\langle 222 \rangle$ 599...604

<400> 127

[illegible]

```

gta gtt ttg ccc ctg gcm agg att gct aca gtt gtg att gga gga gtt      210
Val Val Leu Pro Leu Ala Arg Ile Ala Thr Val Val Ile Gly Gly Val
                               30                               35                               40

gtg gcc atg gcg gct gtg ccc atg gtg ctc agt gcc atg ggc ttc act      258
Val Ala Met Ala Ala Val Pro Met Val Leu Ser Ala Met Gly Phe Thr
                               45                               50                               55

gcg gcg gga atc gcc tcg tcc tcc ata gca gcc aag atg atg tcc gcg      306
Ala Ala Gly Ile Ala Ser Ser Ser Ile Ala Ala Lys Met Met Ser Ala
                               60                               65                               70

gcg gcc att gcc aat sgg ggt gga gtt gcc tcg ggc agc ctt gtg gct      354
Ala Ala Ile Ala Asn Xaa Gly Gly Val Ala Ser Gly Ser Leu Val Ala
                               75                               80                               85

act ctg cag tca ctg gga gca act gga ctc tcc gga ttg acc aag ttc      402
Thr Leu Gln Ser Leu Gly Ala Thr Gly Leu Ser Gly Leu Thr Lys Phe
90                               95                               100                               105

atc ctg ggc tcc att ggg tct gcc att gcg gct gtc att gcg agg ttc      450
Ile Leu Gly Ser Ile Gly Ser Ala Ile Ala Ala Val Ile Ala Arg Phe
                               110                               115                               120

tac tagctccctg cccctcgccc tgcagagaag agaaccatgc caggggagaa      503
Tyr

ggcaccaccagc catcctgacc cagcgaggag ccaactatcc caaatataacc tgggggtgaaa      563
tataccaaat tctgcatctc cagaggaaaa taagaaataa agatgaattg ttacaaaaaa      623
aaaaaaaaaa      633

```

<210> 128
 <211> 588
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 15..236

<220>
 <221> polyA_site
 <222> 573..588

<220>
 <221> polyA_signal
 <222> 551..556

```

<400> 128
gaaaatttttc cact atg gct tcc agc act gtc ccg gtg agc gct gct ggc      50
                               Met Ala Ser Ser Thr Val Pro Val Ser Ala Ala Gly
                               1                               5                               10

tcg gct aat gaa act ccc gaa ata ccg gac aac gtg gga gat tgg ctt      98
Ser Ala Asn Glu Thr Pro Glu Ile Pro Asp Asn Val Gly Asp Trp Leu
                               15                               20                               25

cgg ggc gtc tac cgc ttt gcc act gat agg aat gac ttc cgg agg aac      146
Arg Gly Val Tyr Arg Phe Ala Thr Asp Arg Asn Asp Phe Arg Arg Asn
30                               35                               40

ttg ata cta aat ttg gga ctc ttt gct gcg gga gtt tgg ctg gcc agg      194
Leu Ile Leu Asn Leu Gly Leu Phe Ala Ala Gly Val Trp Leu Ala Arg
45                               50                               55                               60

aac ttg agt gac att gac ctc atg gca cct cag cca ggg gtg      236
Asn Leu Ser Asp Ile Asp Leu Met Ala Pro Gln Pro Gly Val
                               65                               70

tagccaagta gacaaatgga atcctgtgtct gaacccgaat cttccaaaaa acagcctaca      296
atctgtgacc accacaagat gtgccctgat ggcagctgaa gtttgattca gatgggcact      356
tttcttcccc ttccctgcct agtttccttt tgttccttga gtccacgcag aattccattc      416
tctggtcagc agacaggctt aagctaaaagt attgcctcta ttctgtaaag ttctgtacat      476
agttcccaag cttctgcagg gggtgatttt tgctcttgtc ctgagaaata acagtgtgtg      536
tttaaaaaac ataaaataaa taccgcacac aaagacaaaa aaaaaaaaaa aa      588

```

```
<210> 129
<211> 608
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 39..260
```

```
<220>
<221> polyA_site
<222> 589..608
```

```
<220>  
<221> polyA_signal  
<222> 577..582
```

[illegible]

```
<210> 130
<211> 1200
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 29..469
```

```
<220>  
<221> polyA_site  
<222> 1185..1200
```

```

<400> 130
acttggcgag tgagacgctg atgggagg atg gac gta ctg gtg tct gag tgc          52
                               Met Asp Val Leu Val Ser Glu Cys
                               1                               5

tcc gcg cgg ctg ctg cag cag gaa gaa gag att aaa tct ctg act gct          100
Ser Ala Arg Leu Leu Gln Gln Glu Glu Glu Ile Lys Ser Leu Thr Ala
    10                15                20

```

101

gaa att gac cgg ttg aaa aac tgt ggc tgt tta gga gct tct cca aat	148
Glu Ile Asp Arg Leu Lys Asn Cys Gly Cys Leu Gly Ala Ser Pro Asn	
25 30 35 40	
ttg gag cag tta caa gaa gaa aat tta aaa tta aag tat cga ctg aat	196
Leu Glu Gln Leu Gln Glu Glu Asn Leu Lys Leu Lys Tyr Arg Leu Asn	
45 50 55	
att ctt cga aag agt ctt cag gca gaa agg aac aaa cca act aaa aat	244
Ile Leu Arg Lys Ser Leu Gln Ala Glu Arg Asn Lys Pro Thr Lys Asn	
60 65 70	
atg att aac att att agc cgc cta caa gag gtc ttt ggt cat gca att	292
Met Ile Asn Ile Ile Ser Arg Leu Gln Glu Val Phe Gly His Ala Ile	
75 80 85	
aag gct gca tat cca gat ttg gaa aat cct cct ctg cta gtg aca cca	340
Lys Ala Ala Tyr Pro Asp Leu Glu Asn Pro Pro Leu Leu Val Thr Pro	
90 95 100	
agt cag cag gcc aag ttt ggg gac tat cag tgt aat agt gct atg ggt	388
Ser Gln Gln Ala Lys Phe Gly Asp Tyr Gln Cys Asn Ser Ala Met Gly	
105 110 115 120	
att tct cag gtg atg tat tgt cat gac tct tgg ctg ttt gat ttt ttt	436
Ile Ser Gln Val Met Tyr Cys His Asp Ser Trp Leu Phe Asp Phe Phe	
125 130 135	
aag tat tat tat cat cat tgc cat tta cag aaa taatactatt acaagttgta	489
Lys Tyr Tyr Tyr His His Cys His Leu Gln Lys	
140 145	
tccttagtga aaaggacatt tgccacagtt tgaaaaactt gagaaaggag ttgggggggt	549
atatgtttta acttttttag gcacaatttt taaggtttgg ttaaatttta tatgtattct	609
caatatttaa gggcaatcat tgggtactctt ttgttttaggt atttccctcc tgctgtgtcc	669
aggattgctg tgtggtggtg atgagtgcgtg ggaggtgaaa aattaaaata agccatttac	729
cagtcagcat cccaattaaa tatttgatgt aactgtgatc tttgagccag gcttatatat	789
tcattttcaa gcagaggagt tccccatttt aaatagaggc attgtctgat gtgtttatgg	849
ttactgcat ctggcttggtg tctttctgtt ttcctttctt tgctgaattg gaaggggtta	909
ctctgaagag tccaggtcct acagtggtgt ttatttctca agtgtgaata ttgcacacct	969
tcattggcttg aaaattagaa atgtaaatat gctggaacca gacttcataa ctgtagactt	1029
tgttatgtca tcattaaaaa gttcgaggtg ttgtaatccc agctactagg gaggctgagg	1089
caggagaatt gcttgaaccc aggggcggag gttacagtga gccgagatca caccactgca	1149
ctccagcctg ggcaacagag cgagactcca tctctaaaaa aaaaaaaaaa a	1200

<210> 131
 <211> 646
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 50..340

<220>
 <221> polyA_site
 <222> 631..646

<220>
 <221> polyA_signal
 <222> 614..619

<400> 131	
gaggagaaga ggtgaccccg ctaaagataa taaatcatca ttgatcaag atg tct tca	58
Met Ser Ser	
1	
gca cct gag cct cca aca ttc aaa aag gaa cca ccc aaa gaa aaa gag	106
Ala Pro Glu Pro Pro Thr Phe Lys Lys Glu Pro Pro Lys Glu Lys Glu	
5 10 15	
ttt caa agc cca ggg ctc aga ggg gtg cgc aca acc tta ttt cgt	154
Phe Gln Ser Pro Gly Leu Arg Gly Val Arg Thr Thr Thr Leu Phe Arg	

102

20	25	30	35	
gct gtg aat cca gag ctc ttc att aaa cct aac aaa cct gta atg gct				202
Ala Val Asn Pro Glu Leu Phe Ile Lys Pro Asn Lys Pro Val Met Ala				
	40	45	50	
ttc gga ttg gta act ctt tca ctt tgc gtg gca tat att ggt tat cta				250
Phe Gly Leu Val Thr Leu Ser Leu Cys Val Ala Tyr Ile Gly Tyr Leu				
	55	60	65	
cat gca ata caa gag aat aaa aag gac ctc tat gaa gct att gat agt				298
His Ala Ile Gln Glu Asn Lys Lys Asp Leu Tyr Glu Ala Ile Asp Ser				
	70	75	80	
gag ggg cac agt tat atg agg cgg aaa aca tcc aaa tgg gat				340
Glu Gly His Ser Tyr Met Arg Arg Lys Thr Ser Lys Trp Asp				
	85	90	95	
tagtagtgct ggtagtgca gatggacctt tattaaggt tctgaaatct tcaaataaaa				400
gaccttgtag gtgtacagta tcatgtttct tggtctagaa catgctaatag aagagagaag				460
atagcagttg caaccagaca actgtcgtaa attttgtcct ttcacagctg cagccattat				520
ctcattcttt ttccacagag tgagcgtcat aatatattttc tttccttacc tcttataaaa				580
gtgccatgag aatggaagtt gtttttgtag attattaaat tctgtataat aaaaaaaaaa				640
aaaaat				646

<210> 132
 <211> 575
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 177..470

<220>
 <221> polyA_site
 <222> 561..575

<220>
 <221> polyA_signal
 <222> 543..548

<400> 132

atcgaccacg gaaatttgac acctccgggc ttggaagcag ctctctcctc cttccccgct	60
gcttataaac ctcagccctg aggcctccagc tcaactctacc ccactctcctt gccggggtcag	120
ccctgacaaa ggtagcctag ccccttgagg acatcagctt tggcctcagg gtccta atg	179
Met	
1	
gca gca gaa cca ctg aca gag cta gag gag tcc att gag aac gtg gtc	227
Ala Ala Glu Pro Leu Thr Glu Leu Glu Ser Ile Glu Asn Val Val	
	5 10 15
acc acc ttc ttc acc ttt gca agg cag gag ggc cgg aak gat agc ctc	275
Thr Thr Phe Phe Thr Phe Ala Arg Gln Glu Gly Arg Xaa Asp Ser Leu	
	20 25 30
agc gtc aac gag ttc aaa gag ctg gtt acc cag cag ttg ccc cat ctg	323
Ser Val Asn Glu Phe Lys Glu Leu Val Thr Gln Gln Leu Pro His Leu	
	35 40 45
ctc aag gat gtg ggc tct ctt gat gag aag atg aag agc ttg gat gtg	371
Leu Lys Asp Val Gly Ser Leu Asp Glu Lys Met Lys Ser Leu Asp Val	
	50 55 60 65
aat cag gac tcg gag ctc aag ttc aat gag tac tgg aga ttg att ggg	419
Asn Gln Asp Ser Glu Leu Lys Phe Asn Glu Tyr Trp Arg Leu Ile Gly	
	70 75 80
gag ctg gcc aag gaa atc agg aag aag aaa gac ctk aag atc agg aag	467
Glu Leu Ala Lys Glu Ile Arg Lys Lys Lys Asp Leu Lys Ile Arg Lys	
	85 90 95
aag taaagccgcc tggctgagat ggggtgggca gggcagagct gatcagggcc	520
Lys	

gagcagaacc gcactcttcc caaataaagc ttcctccttg aaaaaaaaaa aaaaa 575

<210> 133
 <211> 458
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 75..419

<400> 133
 ctttcctttc tcgttccccg gccatccttag cggctgctgt tgggtggggg ccgtcccgc 60
 cctaaggcag gaag atg gtg gcc gca aag aag acg aaa aag tcg ctg gag 110
 Met Val Ala Ala Lys Lys Thr Lys Lys Ser Leu Glu
 1 5 10
 tcg atc aaa tct agg ctc caa ctc gtt atg aaa agt ggg aag tac gtc 158
 Ser Ile Lys Ser Arg Leu Gln Leu Val Met Lys Ser Gly Lys Tyr Val
 15 20 25
 ctg ggg tac aag cag act ctg aag atg atc aga caa ggc aaa gcg aaa 206
 Leu Gly Tyr Lys Gln Thr Leu Lys Met Ile Arg Gln Gly Lys Ala Lys
 30 35 40
 ttg gtc att ctc gct aac aac tgc cca gct ttg agg aaa tct gaa ata 254
 Leu Val Ile Leu Ala Asn Asn Cys Pro Ala Leu Arg Lys Ser Glu Ile
 45 50 55 60
 gag tac tat gct atg ttg gct aaa act ggt gtc cat cac tac agt ggc 302
 Glu Tyr Tyr Ala Met Leu Ala Lys Thr Gly Val His His Tyr Ser Gly
 65 70 75
 aat aat att gaa ctg ggc aca gca tgc gga aaa tac tac aga gtg tgc 350
 Asn Asn Ile Glu Leu Gly Thr Ala Cys Gly Lys Tyr Tyr Arg Val Cys
 80 85 90
 aca ctg gct atc att gat cca rgt gac tct kac atc att aga agc atg 398
 Thr Leu Ala Ile Ile Asp Pro Xaa Asp Ser Xaa Ile Ile Arg Ser Met
 95 100 105
 cca gaa cag act ggt gaa aag taaacctttt cacctacaaa atttcacctg 449
 Pro Glu Gln Thr Gly Glu Lys
 110 115
 caaacctta 458

<210> 134
 <211> 588
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 144..449

<220>
 <221> polyA_site
 <222> 573..588

<220>
 <221> polyA_signal
 <222> 552..557

<400> 134
 gaggggagcag ggggctagcg ggcacggctg gggcgagcgc gcctgcgcgc tgggtgattt 60
 tttcacgtgt cgccagggcc ggactgcgag tctctttgcg gcgctacact agagcagagt 120
 acgagtctga ggcgaggga gta atg gca gga caa gcg ttt aga aag ttt ctt 173
 Met Ala Gly Gln Ala Phe Arg Lys Phe Leu
 1 5 10
 cca ctc ttt gac cga gta ttg gtt gaa agg agt cct gct gaa act gta 221

104

```

Pro Leu Phe Asp Arg Val Leu Val Glu Arg Ser Pro Ala Glu Thr Val
      15      20      25
acc aaa gga ggc att atg ctt cca gaa aaa tct caa gga aaa gta ttg      269
Thr Lys Gly Gly Ile Met Leu Pro Glu Lys Ser Gln Gly Lys Val Leu
      30      35      40
caa gca aca gta gtc gct gtt gga tcg ggt tct aaa gga aag ggt gga      317
Gln Ala Thr Val Val Ala Val Gly Ser Gly Ser Lys Gly Lys Gly Gly
      45      50      55
gag att caa cca gtt agc gtg aaa gtt gga gat aaa gtt ctt ctc cca      365
Glu Ile Gln Pro Val Ser Val Lys Val Gly Asp Lys Val Leu Leu Pro
      60      65      70
gaa tat gga ggc acc aaa gta gtt yta gat aac aag gat tat ttc cta      413
Glu Tyr Gly Gly Thr Lys Val Val Leu Asp Asn Lys Asp Tyr Phe Leu
      75      80      85      90
ttt aga gat ggt gac att ctt gga aag tac gta gac tgaaataagt      459
Phe Arg Asp Gly Asp Ile Leu Gly Lys Tyr Val Asp
      95      100
cactattgaa atggcatcaa catgatgctg cccattccac tgaagttctg aaatctttcg      519
tcatgtaa atttccata tttctctttt ataataaact aatgataact aacaaaaaaaa      579
aaaaaaaaa      588

<210> 135
<211> 538
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 93..398

<220>
<221> polyA_site
<222> 523..538

<220>
<221> polyA_signal
<222> 501..506

<400> 135
gggtgatttt ttcacgtgtc gccagggccg gactgcgagt ctctttgcgg cgctacacta      60
gagcagagta cgagtctgag gcggagggag ta atg gca gga caa gcg ttt aga      113
                               Met Ala Gly Gln Ala Phe Arg
                               1      5
aag ttt ctt cca ctc ttt gac cga gta ttg gtt gaa agg agt gct gct      161
Lys Phe Leu Pro Leu Phe Asp Arg Val Leu Val Glu Arg Ser Ala Ala
      10      15      20
gaa act gta acc aaa gga ggc att atg ctt cca gaa aaa tct caa gga      209
Glu Thr Val Thr Lys Gly Gly Ile Met Leu Pro Glu Lys Ser Gln Gly
      25      30      35
aaa gta ttg caa gca aca gta gtc gct gtt gga tcg ggt tct aaa gga      257
Lys Val Leu Gln Ala Thr Val Val Ala Val Gly Ser Gly Ser Lys Gly
      40      45      50      55
aag ggt gga gag att caa cca gtt agc gtg aaa gtt gga gat aaa gtt      305
Lys Gly Gly Glu Ile Gln Pro Val Ser Val Lys Val Gly Asp Lys Val
      60      65      70
ctt ctc cca gaa tat gga ggc acc aaa gta gtt cta gat gac aag gat      353
Leu Leu Pro Glu Tyr Gly Gly Thr Lys Val Val Leu Asp Asp Lys Asp
      75      80      85
tat ttc cta ttt aga gat ggt gac att ctt gga aag tac gta gac      398
Tyr Phe Leu Phe Arg Asp Gly Asp Ile Leu Gly Lys Tyr Val Asp
      90      95      100
tgaaataagt cactattgaa atggcatcaa catgatgctg cccattccac tgaagttctg      458
aaatctttcg tcatgtaa atttccata tttctctttt ataataaact aatgataact      518

```

aacgaaaaaa aaaaaaaaaa

538

<210> 136
 <211> 912
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 97..444

<220>
 <221> polyA_site
 <222> 897..912

<400> 136
 atctttccgt cccgggcagc cagcgccagt cggagccagc gcgagccgcc gccgccatca 60
 ctgccgctgc caagtccctcc acccgctgcc cccgcc atg tct gct acc gct gcc 114
 Met Ser Ala Thr Ala Ala
 1 5
 acg gcc ccc cct gct gcc ccg gct ggg gag ggt ggt ccc cct gca ccc 162
 Thr Ala Pro Pro Ala Ala Pro Ala Gly Glu Gly Gly Pro Pro Ala Pro
 10 15 20
 cct cca aac ctg acc agt aac agg aga ctg cag cag acc cag gcc cag 210
 Pro Pro Asn Leu Thr Ser Asn Arg Arg Leu Gln Gln Thr Gln Ala Gln
 25 30 35
 gtg gat gag gtg gtg gac atc atg agg gtg aac gtg gac aag gtc ctg 258
 Val Asp Glu Val Val Asp Ile Met Arg Val Asn Val Asp Lys Val Leu
 40 45 50
 gag cga gac cag aag ctg tcg gag ctg gac gac cgt gca gat gca ctg 306
 Glu Arg Asp Gln Lys Leu Ser Glu Leu Asp Asp Arg Ala Asp Ala Leu
 55 60 65 70
 cag gcg ggg ccc tcc cag ttt gaa aca agc gca gcc aag ctg aag cgc 354
 Gln Ala Gly Pro Ser Gln Phe Glu Thr Ser Ala Ala Lys Leu Lys Arg
 75 80 85
 aaa tac tgg tgg aaa aac ctg aag atg atg atc atc ttg gga gtg att 402
 Lys Tyr Trp Trp Lys Asn Leu Lys Met Met Ile Ile Leu Gly Val Ile
 90 95 100
 tgc gcc atc atc ctg atc atc atc ata gtt tac ttc agc act 444
 Cys Ala Ile Ile Leu Ile Ile Ile Ile Val Tyr Phe Ser Thr
 105 110 115
 taaatccccg aggagtctgc cctgcctaga gaagggcctc tcccccaacc ctgagccgtt 504
 cctccacctc tcagccatat ctttcagccc cccctcccct ggatccgtgt gtgtgtgtgt 564
 ccgtgtgtgt gtccccctgt aaatagccag ctgttattta tacatatata atattatata 624
 tatttggtct gttttagtct ttattactag atgatttttc cggttgtcct taacaccctt 684
 tcctgaggtt cccttcaccc ctctctcttg ccttccttcc ctttcccttt cttcctgact 744
 agccccaagt cccttcattt gcattctgcta tgcaatagtc cctctccttt ccttcttctt 804
 ccctcagatt tagctgatcc ttcctccac cctggccttc ctttctcttt tcctcctcac 864
 tctccccgtc atgctccctc tgccccgccc tcaaaaaaaaa aaaaaaaa 912

<210> 137
 <211> 562
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 45..335

<220>
 <221> polyA_site
 <222> 545..562

<220>

<221> polyA_signal

<222> 526..531

<400> 137

```

tggtattttg tagctttaat gattttttggt ttctaataagg gcaa atg tct cta agc      56
                                   Met Ser Leu Ser
                                   1
ttg gtg ttt aga gct gct tca tat ttt aaa cta gtt cca ttc cac agt      104
Leu Val Phe Arg Ala Ala Ser Tyr Phe Lys Leu Val Pro Phe His Ser
5                               10                               15                               20
tct agt tca aac cag ttt tta cag cct cct ggg tgg gtc gtc ttg acc      152
Ser Ser Ser Asn Gln Phe Leu Gln Pro Pro Gly Trp Val Val Leu Thr
25                               30                               35
caa act ctt gtg ttg tta cat ttt gag agg ttt tca tac cag aat gta      200
Gln Thr Leu Val Leu Leu His Phe Glu Arg Phe Ser Tyr Gln Asn Val
40                               45                               50
cca aaa agt gca caa ggt aaa ggt aat tta cag cca gaa aca aat ata      248
Pro Lys Ser Ala Gln Gly Lys Gly Asn Leu Gln Pro Glu Thr Asn Ile
55                               60                               65
cat ttg ttt cat ttc ttg act ttc cct aag cag ata agc aga aac ttg      296
His Leu Phe His Phe Leu Thr Phe Pro Lys Gln Ile Ser Arg Asn Leu
70                               75                               80
ttt aac tca tta ctt tgt ttg atg tgt ctt aca tat ttt tgactaaaag      345
Phe Asn Ser Leu Leu Cys Leu Met Cys Leu Thr Tyr Phe
85                               90                               95
ttatagaatg ttattcctct gggggaatct ttacaggtg gaggaatggg gatagcagta      405
ttgcctcaga ttcaaactgg catcaccata aaccctgtag gccarggtgg aatgaagtca      465
gctccttttt atagttgaaa tacaattttt tattcaccat tgtctgcaca aatccttgaa      525
aataaacctt tttttcccta caaaaaaaaa aaaaaaa      562

```

<210> 138

<211> 566

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 32..361

<220>

<221> polyA_site

<222> 551..566

<220>

<221> polyA_signal

<222> 532..537

<400> 138

```

aacttgctct gcccgctcaa acagttgcag g atg tcg atg aca gac ttg ctg      52
                                   Met Ser Met Thr Asp Leu Leu
                                   1                               5
aac gct gag gac atc aag aag gcg gtg gga gcc ttt agc gct acc gac      100
Asn Ala Glu Asp Ile Lys Lys Ala Val Gly Ala Phe Ser Ala Thr Asp
10                               15                               20
tcc ttc gac cac aaa aag ttc ttc caa atg gtc ggc ctg aag aaa aag      148
Ser Phe Asp His Lys Lys Phe Phe Gln Met Val Gly Leu Lys Lys Lys
25                               30                               35
agt gcg gat gat gtg aag aag gtg ttt cac atg ctg gac aag gac aaa      196
Ser Ala Asp Asp Val Lys Lys Val Phe His Met Leu Asp Lys Asp Lys
40                               45                               50                               55
agt ggc ttc atc gag gag gat gag ctg gga ttc atc cta aaa ggc ttc      244
Ser Gly Phe Ile Glu Glu Asp Glu Leu Gly Phe Ile Leu Lys Gly Phe

```

107

	60		65		70	
tcc cca gat gcc aga gac ctg tct gct aaa gaa acc aag atg ctg atg						292
Ser Pro Asp Ala Arg Asp Leu Ser Ala Lys Glu Thr Lys Met Leu Met						
	75		80		85	
gct gct gga gac aaa gat ggg gac ggc aaa att ggg gtt gac gaa ttc						340
Ala Ala Gly Asp Lys Asp Gly Asp Gly Lys Ile Gly Val Asp Glu Phe						
	90		95		100	
tcc act ctg gtg gct gaa agc taagaagcac tgactgcccc tggctcttcca						391
Ser Thr Leu Val Ala Glu Ser						
	105		110			
cctctctgcc ctcaacacccc aatctcggcc cctcttgcca ccctcctgca tttctgttca						451
gttcgtttat gttatTTTTT actcccccat cccctgtggc cctctaataga caccattctt						511
ctggaaaatg ctggagaagc aataaagggt gtaccagtca aaaaaaaaaa aaaaaa						566

<210> 139
 <211> 567
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 31..360

<220>
 <221> polyA_site
 <222> 552..567

<220>
 <221> polyA_signal
 <222> 531..536

<400> 139	
acttgctctg cccgctcaaaa cagttgcagg atg tcg atg aca gac ttg ctg aac	54
	Met Ser Met Thr Asp Leu Leu Asn
	1 5
gct gag gac atc aag aag gcg gtg gga gcc ttt agc gct acc gac tcc	102
Ala Glu Asp Ile Lys Lys Ala Val Gly Ala Phe Ser Ala Thr Asp Ser	
	10 15 20
ttc gac cac aaa aag ttc ttc caa atg gtc ggc ctg aag aaa aag agt	150
Phe Asp His Lys Lys Phe Phe Gln Met Val Gly Leu Lys Lys Lys Ser	
	25 30 35 40
gcg gat gat gtg aag aag gtg ttt cac atg ctg gac aag gac aaa agt	198
Ala Asp Asp Val Lys Lys Val Phe His Met Leu Asp Lys Asp Lys Ser	
	45 50 55
ggc ttc atc gag gag gat gag ctg gga ttc atc cta aaa ggc ttc tcc	246
Gly Phe Ile Glu Glu Asp Glu Leu Gly Phe Ile Leu Lys Gly Phe Ser	
	60 65 70
cca gat gcc aga gac ctg tct gct aaa gaa acc aag atg ctg atg gct	294
Pro Asp Ala Arg Asp Leu Ser Ala Lys Glu Thr Lys Met Leu Met Ala	
	75 80 85
gct gga gac aaa gat ggg gac ggc aaa att ggg gtt gac gaa ttc tcc	342
Ala Gly Asp Lys Asp Gly Asp Gly Lys Ile Gly Val Asp Glu Phe Ser	
	90 95 100
act ctg gtg gct gaa agc taagaagcac tgactgcccc tggctcttcca	390
Thr Leu Val Ala Glu Ser	
	105 110
cctctctgcc ctgaacacccc aatctcggcc cctctcgcca ccctcctgca tttctgttca	450
gttcgtttat gttatTTTTT actcccccat cccctgtggc cctctaataga caccattctt	510
ctggaaaatg ctggagaagc aataaagggt gtaccagtca aaaaaaaaaa aaaaaa	567

<210> 140
 <211> 566
 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 32..361

<220>

<221> polyA_site

<222> 551..566

<220>

<221> polyA_signal

<222> 532..537

<400> 140

```

aacttgctct gcccgctcaa acagttgcag g atg tcg atg aca gac ttg ctg      52
                               Met Ser Met Thr Asp Leu Leu
                               1           5
aac gct gag gac atc aag aag gcg gtg gga gcc ttt agc gct acc gac      100
Asn Ala Glu Asp Ile Lys Lys Ala Val Gly Ala Phe Ser Ala Thr Asp
          10           15           20
tcc ttc gac cac aaa aag ttc ttc caa atg gtc ggc ctg aag aaa aag      148
Ser Phe Asp His Lys Lys Phe Phe Gln Met Val Gly Leu Lys Lys Lys
          25           30           35
agt gcg gat gat gtg aag aag gtg ttt cac atg ctg gac aag gac aaa      196
Ser Ala Asp Asp Val Lys Lys Val Phe His Met Leu Asp Lys Asp Lys
          40           45           50           55
agt ggc ttc atc gag gag gat gag ctg gga ttc atc cta aaa ggc ttc      244
Ser Gly Phe Ile Glu Glu Asp Glu Leu Gly Phe Ile Leu Lys Gly Phe
          60           65           70
tcc cca gat gcc aga gac ctg tct gct aaa gaa acc aag atg ctg atg      292
Ser Pro Asp Ala Arg Asp Leu Ser Ala Lys Glu Thr Lys Met Leu Met
          75           80           85
gct gct gga gac aaa gat ggg gac ggc aaa att ggg gtt gac gaa ttc      340
Ala Ala Gly Asp Lys Asp Gly Asp Gly Lys Ile Gly Val Asp Glu Phe
          90           95           100
tcc act ctg gtg gct gaa agc taagaagcac tgactgcccc tggtcttcca      391
Ser Thr Leu Val Ala Glu Ser
          105           110
cctctctgcc ctcaacacccc aatctcggcc cctcttgcca ccctcctgca tttctgttca      451
gttcgtttat gttatttttt actcccccac cccctgtggc cctctaataga caccattctt      511
ctggaaaatg ctggagaagc aataaagggt gtaccagtca aaaaaaaaaa aaaaa      566

```

<210> 141

<211> 625

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 85..420

<220>

<221> polyA_site

<222> 610..625

<400> 141

```

cccttctgcg cggtcacgcc gagccagcgc ctgggcctgg aaccggggccg tagccccccc      60
agtttcgccc accacctccc tacc atg gac ccc cgc aaa gtg aac gak ctt      111
                               Met Asp Pro Arg Lys Val Asn Xaa Leu
                               1           5
cgg gcc ttt gtg aaa atg tgt aag cag gat ccg agc gtt ctg cac acc      159
Arg Ala Phe Val Lys Met Cys Lys Gln Asp Pro Ser Val Leu His Thr

```

109																						
10											15						20				25	
gag gaa atg cgc ttc	ctg agg gas tgg	gtg gag agc atg ggt	ggt aaa	207																		
Glu Glu Met Arg Phe	Leu Arg Xaa Trp	Val Glu Ser Met Gly	Gly Lys																			
30				35	40																	
gta cca cct gct act	caa aaa gct aaa	tca gaa gaa aat acc	aag gaa	255																		
Val Pro Pro Ala Thr	Gln Lys Ala Lys	Ser Glu Glu Asn Thr	Lys Glu																			
45				50	55																	
gaa aaa cct gat agt	aag aag gtg gag	gaa gac tta aag	gca gac gaa	303																		
Glu Lys Pro Asp Ser	Lys Lys Val Glu	Glu Asp Leu Lys	Ala Asp Glu																			
60				65	70																	
cca tca agt gag gaa	agt gat cta gaa	att gat aaa gaa	ggt gtg att	351																		
Pro Ser Ser Glu Glu	Ser Asp Leu Glu	Ile Asp Lys Glu	Gly Val Ile																			
75				80	85																	
gaa cca gac act gat	gct cct caa gaa	atg gga gat gaa	aat gcg gag	399																		
Glu Pro Asp Thr Asp	Ala Pro Gln Glu	Met Gly Asp Glu	Asn Ala Glu																			
90				95	100																	
ata acg gct caa cat	ttt ttg tagccacagg	cttccacgga cttcacgtca		450																		
Ile Thr Ala Gln His	Phe Leu																					
110																						
ttattggctc aactttcctc	actatctgct tcatccgcca	actaatatatt cactttacat		510																		
ccaaacatca ctttggttc	gaagccgccg cctgatactg	gcattttgta gatgtggttt		570																		
gactatttct gtatgtctcc	atctattgat gaggggtctta	aaaaaaaaa aaaat		625																		
<p><210> 142</p> <p><211> 685</p> <p><212> DNA</p> <p><213> Homo sapiens</p> <p><220></p> <p><221> CDS</p> <p><222> 22..375</p> <p><220></p> <p><221> polyA_site</p> <p><222> 670..685</p> <p><220></p> <p><221> polyA_signal</p> <p><222> 650..655</p> <p><400> 142</p>																						
agtcgagcca gcgtcgccgc	g atg gtg ttg ttg	gag agc gag cag	ttc ctg	51																		
	Met Val Leu Leu	Glu Ser Glu Gln	Phe Leu																			
1				5	10																	
acg gag ctg acc aga	ctt ttc cag aag	tgc cgg acg tcg	ggc agc gtc	99																		
Thr Glu Leu Thr Arg	Leu Phe Gln Lys	Cys Arg Thr Ser	Gly Ser Val																			
15				20	25																	
tat atc acc ttg aag	aag tat gac ggt	cga acc aaa ccc	att cca aag	147																		
Tyr Ile Thr Leu Lys	Lys Tyr Asp Gly	Arg Thr Lys Pro	Ile Pro Lys																			
30				35	40																	
aaa ggt act gtg gag	ggc ttt gag ccc	gca gac aac aag	tgt ctg tta	195																		
Lys Gly Thr Val Glu	Gly Phe Glu Pro	Ala Asp Asn Lys	Cys Leu Leu																			
45				50	55																	
aga gct acc gat ggg	aag aag aag atc	agc act gtg gtg	agc tcc aag	243																		
Arg Ala Thr Asp Gly	Lys Lys Lys Ile	Ser Thr Val Val	Ser Ser Lys																			
60				65	70																	
gaa gtg aat aag ttt	cag atg gct tat	tca aac ctc ctt	aga gct aac	291																		
Glu Val Asn Lys Phe	Gln Met Ala Tyr	Ser Asn Leu Leu	Arg Ala Asn																			
75				80	85																	
atg gat ggg ttg aag	aag aga gac aaa	aag aac gaa act	aag aag acc	339																		
Met Asp Gly Leu Lys	Lys Arg Asp Lys	Lys Asn Glu Thr	Lys Lys Thr																			
95				100	105																	
aaa gca gca gca gca	gca gca gca gca	aca gca gca cag	taaagggcat	385																		

110

Lys Ala Ala Ala Ala Ala Ala Ala Thr Ala Ala Gln

	110	115	
acatttcctg	ctttcaccaa	ttaaccactg	aattgctatt
ctaggtttct	ggttcccca	cagtaggtgt	tttcacataa
gaatagttgc	agtggtttata	ggatagttgt	ggtaagaatc
taattggtct	gtgctgcatg	gttatatact	cctggattat
catctctgtg	aagagcaagc	tatcattaaa	catgtctgtt
			tatcaaaaaa
			aaaaaaaaa

445
505
565
625
685

<210> 143
<211> 553
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 9..335

<220>
<221> polyA_site
<222> 538..553

<220>
<221> polyA_signal
<222> 520..525

<400> 143	
tcaagaat atg gat agt tta cct tct ggt aaa ata cat cga aaa gtg aaa	50
Met Asp Ser Leu Pro Ser Gly Lys Ile His Arg Lys Val Lys	
1 5 10	
ata ata tta gga cga aat aga aaa gaa aat ctg gaa cca aat gct gaa	98
Ile Ile Leu Gly Arg Asn Arg Lys Glu Asn Leu Glu Pro Asn Ala Glu	
15 20 25 30	
ttt gat aaa aga act gaa ttt att aca caa gaa gaa aac aga att tgt	146
Phe Asp Lys Arg Thr Glu Phe Ile Thr Gln Glu Glu Asn Arg Ile Cys	
35 40 45	
agt tca ccg gta cag tct tta cta gac ttg ttt cag act agt gaa gag	194
Ser Ser Pro Val Gln Ser Leu Leu Asp Leu Phe Gln Thr Ser Glu Glu	
50 55 60	
aaa tca gaa ttt ttg ggt ttc aca agc tac aca gaa aag agt ggt ata	242
Lys Ser Glu Phe Leu Gly Phe Thr Ser Tyr Thr Glu Lys Ser Gly Ile	
65 70 75	
tgc aat gtt tta gat att tgg gaa gag gaa aat tca gat aat ctg tta	290
Cys Asn Val Leu Asp Ile Trp Glu Glu Glu Asn Ser Asp Asn Leu Leu	
80 85 90	
aca gcg ttt ttc tcg tcc cct tca act tct aca ttt act ggc ttt	335
Thr Ala Phe Phe Ser Ser Pro Ser Thr Ser Thr Phe Thr Gly Phe	
95 100 105	
tagaatttaa aaaatgcata cttttcagaa gtgataagga tcatattcctt gaaattttta	395
taaatatgta tggaaattct taggattttt ttaccagctt tgtttacaga cccaaatgta	455
aatattaaaa ataaatattt gcaattttct acagaattga atacctgtta aagaaaaaatt	515
acagaataaa cttgtgactg gtaaaaaaaaa aaaaaaat	553

<210> 144
<211> 631
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 156..464

<220>
<221> polyA_site

<222> 610..631

<220>

<221> polyA_signal

<222> 559..564

<400> 144

```

aattcatatc tgagtgggcg gtggcgattg gtgttggcgg tctggctcag ctgggcaggg      60
ggtaacttta ctgatttggg ggtggttttt agtttaattt ttcttttcta gcttcccatc    120
gacggtcagt gcgcacgttg taatcagctg aggcc atg tca gga gac gga gcc      173
                               Met Ser Gly Asp Gly Ala
                               1             5

acg gag cag gca gct gag tat gtc cca gag aag gtg aag aaa gcg gaa      221
Thr Glu Gln Ala Ala Glu Tyr Val Pro Glu Lys Val Lys Lys Ala Glu
                               10             15             20

aag aaa tta gaa gag aat cca tat gac ctt gat gct tgg agc att ctc      269
Lys Lys Leu Glu Glu Asn Pro Tyr Asp Leu Asp Ala Trp Ser Ile Leu
                               25             30             35

att cga gag gca cag aat caa cct ata gac aaa gca cgg aag act tat      317
Ile Arg Glu Ala Gln Asn Gln Pro Ile Asp Lys Ala Arg Lys Thr Tyr
                               40             45             50

gaa cgc ctt gtt gcc cag ttc ccc agt tct ggc aga ttc tgg aaa ctg      365
Glu Arg Leu Val Ala Gln Phe Pro Ser Ser Gly Arg Phe Trp Lys Leu
55             60             65             70

tac att gaa gca gag gtt act att tta ttt tat ttt ttc tta tat cag      413
Tyr Ile Glu Ala Glu Val Thr Ile Leu Phe Tyr Phe Phe Leu Tyr Gln
                               75             80             85

tat tgc agc att cac tgt agt gat aga aaa caa gtt asg aac ata gcc      461
Tyr Cys Ser Ile His Cys Ser Asp Arg Lys Gln Val Xaa Asn Ile Ala
                               90             95             100

mat taggacaagg aggatttaaa tgtgtcttac ctttattttg taaaataggt      514
Xaa

ataaaggagt aattaataatg aatttttgaa atttgggtct tttacaagct gatgattgtt    574
gcattttgga gttgcaacaa cattaataaca gtttcaatga taaaaaaaaa aaaaaaa      631

```

<210> 145

<211> 635

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 160..375

<220>

<221> polyA_site

<222> 620..635

<220>

<221> polyA_signal

<222> 569..574

<400> 145

```

ttttaattca tatctgagtg ggcggtggcg attggtgttg gcggtctggc tcagctgggc      60
agggggtaac tttactgatt tgggggtggt ttttagttta atttttcttt tctagcttcc    120
catcgacggt cagtgcgcac gttgtaatca gctgaggcc atg tca gga gac gga      174
                               Met Ser Gly Asp Gly
                               1             5

gcc acg gag cag gca gct gag tat gtc tca gag aag gtg aag aaa gcg      222
Ala Thr Glu Gln Ala Ala Glu Tyr Val Ser Glu Lys Val Lys Lys Ala
                               10             15             20

gaa aag aaa tta gaa gag aat cca tat gac ctt gat gct tgg agc att      270
Glu Lys Lys Leu Glu Glu Asn Pro Tyr Asp Leu Asp Ala Trp Ser Ile

```


112

```

      25              30              35
ctc att cga gag gca cag aat caa cct ata gac aaa gca cgg aag act      318
Leu Ile Arg Glu Ala Gln Asn Gln Pro Ile Asp Lys Ala Arg Lys Thr
      40              45              50
tat gaa cgc ctt gtt gcc cag ttc ccc agt tct ggc aga ttc tgg aaa      366
Tyr Glu Arg Leu Val Ala Gln Phe Pro Ser Ser Gly Arg Phe Trp Lys
      55              60              65
ctg tac att taagcagagg ttactatattt attttatattt ttcttatatc      415
Leu Tyr Ile
70
agtattgcag cattcactgt agtgatagaa aamaagttas gaacatagcc aattaggaca      475
aggaggattt aaatttgtct tacctttatt ttgtaaaata ggtataaagg agtaattaaa      535
atgaattttt gaaatttggg tcttttacaa gctgatgatt gttgcatttt ggagttgcaa      595
caacattaaa acagtttcaa tggcaaaaaa aaaaaaaaaa      635

<210> 146
<211> 733
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 32..349

<220>
<221> polyA_site
<222> 718..733

<220>
<221> polyA_signal
<222> 711..716

<400> 146
actaactagc aaacgggggac tagaaatagg g atg ctg aaa agc aac ggg gag      52
                               Met Leu Lys Ser Asn Gly Glu
                               1              5
aga cgc agt cgt aac gca ctt ccg gcg gtc tac gcg agg aag atg gct      100
Arg Arg Ser Arg Asn Ala Leu Pro Ala Val Tyr Ala Arg Lys Met Ala
      10              15              20
gca tcc cag cag caa gct tca gcg gct tcc tca gct gct ggt gta tcg      148
Ala Ser Gln Gln Gln Ala Ser Ala Ala Ser Ser Ala Ala Gly Val Ser
      25              30              35
ggt cct agt tcg gct ggc ggc ccg ggt ccc cag cag ccg caa ccg      196
Gly Pro Ser Ser Ala Gly Gly Pro Gly Pro Gln Gln Pro Gln Pro
      40              45              50              55
cca gca caa ctg gtg ggc cct gcc cag agc ggc ctc ctg cag caa cag      244
Pro Ala Gln Leu Val Gly Pro Ala Gln Ser Gly Leu Leu Gln Gln Gln
      60              65              70
caa cag gac ttc gat cct gtg cag cgt tat aag atg ctc atc ccg cag      292
Gln Gln Asp Phe Asp Pro Val Gln Arg Tyr Lys Met Leu Ile Pro Gln
      75              80              85
ctg aag gag agt cta cag gtg att ggc ctt aag cag cga gaa gca aac      340
Leu Lys Glu Ser Leu Gln Val Ile Gly Leu Lys Gln Arg Glu Ala Asn
      90              95              100
tgg att tgg tagtctagag gggcaggccg rgggccaggt tagtcggaga      389
Trp Ile Trp
      105
gagcgcctaag cagaaagagg aatagaacct gctgtttttgg cctggctggt gcacgcctgt      449
gttcccagca ctttgggagg ccgaggcggg cggatcactt gaggtcagga gttcaagacc      509
agcctggtga aacctcgtct ctactaaaaa tacaaaaatt agccggggcat ggtggcgcg      569
gcctgtaatt acagctactc gggaagctga ggcaggagaa tcgcttgaac ccaggaggcg      629
gaggttgacg tgagccgaca tcgtgccatt gcactccagc ctgggcaaca gagcagagtc      689
cgtcaawaaa ataaataaat aaataaataa aaaaaaaaaa aaaa      733

```

<210> 147
 <211> 521
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 111..368

<220>
 <221> polyA_site
 <222> 506..521

<220>
 <221> polyA_signal
 <222> 483..488

<400> 147
 tctcagcttc cggctggtag tagttccgct tcctgtccga ctgtggtgtc tttgctgagg 60
 gtcacattga gctgcagggt gaatccgggg tgcctttagg attcagcacc atg gcg 116
 Met Ala
 1
 gaa gac atg gag acc aaa atc aag aac tac aag acc gcc cct ttt gac 164
 Glu Asp Met Glu Thr Lys Ile Lys Asn Tyr Lys Thr Ala Pro Phe Asp
 5 10 15
 agc cgc ttc ccc aac cag aac cag act aga aac tgc tgg cag aac tac 212
 Ser Arg Phe Pro Asn Gln Asn Gln Thr Arg Asn Cys Trp Gln Asn Tyr
 20 25 30
 ctg gac ttc cac cgc tgt cag aag gca atg acc gct aaa gga ggc gat 260
 Leu Asp Phe His Arg Cys Gln Lys Ala Met Thr Ala Lys Gly Gly Asp
 35 40 45 50
 atc tct gtg tgc gaa tgg tac cag cgt gtg tac cag tcc ctc tgc ccc 308
 Ile Ser Val Cys Glu Trp Tyr Gln Arg Val Tyr Gln Ser Leu Cys Pro
 55 60 65
 aca tcc tgg gtc aca gac tgg gat gag caa cgg gct gaa ggc acg ttt 356
 Thr Ser Trp Val Thr Asp Trp Asp Glu Gln Arg Ala Glu Gly Thr Phe
 70 75 80
 ccc ggg aag aty tgaactggct gcattctcct ttctctgtc ctccatcctt 408
 Pro Gly Lys Ile
 85
 ctcccaggat ggtgaagggg gacctggtac ccagtgatcc ccaccccagg atcctaaatc 468
 atgacttacc tgctaataaa aactcattgg aaaagtcaaa aaaaaaaaaa aaa 521

<210> 148
 <211> 547
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 95..409

<220>
 <221> polyA_site
 <222> 533..547

<220>
 <221> polyA_signal
 <222> 509..514

<400> 148
 ttcgaagtgg ttgtctctgc actatttact gaaaagttca tcctttcccc agtgatttgt 60

114

```

aacactgcct ctttcataaa ttaaattttt gtgt atg tgt gtg ggt ctt ttg atg      115
Met Cys Val Gly Leu Leu Met
1      5
gtt tct att ctg act gac atc aat ttg tct aat ctt gta gca gta cag      163
Val Ser Ile Leu Thr Asp Ile Asn Leu Ser Asn Leu Val Ala Val Gln
10      15      20
tac agt cct gat tat tgc aac ttt agg aaa agg tct gat aaa aac caa      211
Tyr Ser Pro Asp Tyr Cys Asn Phe Arg Lys Arg Ser Asp Lys Asn Gln
25      30      35
gat gcc tcc aca ttt tgt cat aat tgt aac caa ttt cac ctc tgt ctc      259
Asp Ala Ser Thr Phe Cys His Asn Cys Asn Gln Phe His Leu Cys Leu
40      45      50      55
cag tat cac cac aaa att gtt ctt cct tgg agt gtc ttg gct att ctt      307
Gln Tyr His His Lys Ile Val Leu Pro Trp Ser Val Leu Ala Ile Leu
60      65      70
agc caa ctg ttc ctc cat att act tct aga att aga cca aca att tat      355
Ser Gln Leu Phe Leu His Ile Thr Ser Arg Ile Arg Pro Thr Ile Tyr
75      80      85
aaa tca aac aaa ccc aag agc att gaa att ttg att gga ttt gta ttg      403
Lys Ser Asn Lys Pro Lys Ser Ile Glu Ile Leu Ile Gly Phe Val Leu
90      95      100
aat tta tagattaatc tggtaaacca tgtcatcttt acaatgttgt cttccaatac      459
Asn Leu
105
atgaatatgg tacagctctt catttactta ggcctttgaa atatctttca ataaagttaa      519
taattttctc cataaaaaaa aaaaaaaa      547

```

<210> 149

<211> 3159

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 103..405

<400> 149

```

acttccggtg cgaaccgcct cggccgttcc ctccgggaact tactgagcgc ggccgccgag      60
cccagctccg ccgccgagcg cctgtgccgg cacggctaca cc atg gag cgc ccg      114
Met Glu Arg Pro
1
gat aag gcg gcg ctg aac gca ctg cag cct cct gag ttc aga aat gaa      162
Asp Lys Ala Ala Leu Asn Ala Leu Gln Pro Pro Glu Phe Arg Asn Glu
5      10      15      20
agc tca tta gca tct aca ctg aag acg ctc ctg ttc ttc aca gct tta      210
Ser Ser Leu Ala Ser Thr Leu Lys Thr Leu Leu Phe Phe Thr Ala Leu
25      30      35
atg atc act gtt cct att ggg tta tat ttc aca act aaa tct tac ata      258
Met Ile Thr Val Pro Ile Gly Leu Tyr Phe Thr Thr Lys Ser Tyr Ile
40      45      50
ttt gaa ggc gcc ctt ggg atg tcc aat agg gac agc tat ttt tac gct      306
Phe Glu Gly Ala Leu Gly Met Ser Asn Arg Asp Ser Tyr Phe Tyr Ala
55      60      65
gct att gtt gca gtg gtc gcc gtc cat gtg gtg ctg gcc ctc ttt gtg      354
Ala Ile Val Ala Val Val Ala Val His Val Val Leu Ala Leu Phe Val
70      75      80
tat gtg gcc tgg aat gaa ggc tca cga cag tgg cgt gaa ggc aaa cag      402
Tyr Val Ala Trp Asn Glu Gly Ser Arg Gln Trp Arg Glu Gly Lys Gln
85      90      95      100
gat taaagtgaac atcacctttt tatagcatta aattcatttt ttaaaatgat      455
Asp
aaatgctgga gggggccatc tgatttgaat aaagttgaaa gaacatgtta aagtcagtct      515
taaggagtca cgtttgagta tgtaaatttt gatctttcta atatgttggt ttgtatatcc      575

```

```

agttttaact gtatgaatct gatttgcaaa tgagaatttg gaaaagttag ttacaaagaa 635
atatgttaat ttaattagac aatactctgg aaggaatttt atcttctttc aacaaaacat 695
gttttatagt attctgactt acggttgctt ttgagtttta ctcatktgga tatattaaga 755
tgcacacagt gaagcaaatt aaactccact ttacgctgga atgctttctt tagcatgaaa 815
ataccaggtc cttggatttg ggattttaat ttcctatgga aagttgctta aattgtggac 875
actggaatta atctgaatgt cactgaggaa tttcacatga agtgtaatcc ctagtcaata 935
agaattatcc attacattat tttatgggaa aactaggcta aattacatcc attcaggtaa 995
aaggaccta gcttactgaa ggatctaaag agcaaagcaa agatctcact actcaaacac 1055
tcagcctgct tctttcaagt ccccttgcac gccagcttag tgctttgcag accaactttt 1115
taatgagata ctttgcttcc tcattcaaca ttgaagctag gcttcaatta aaaggttcga 1175
ggaagctcca tttaaaattg tttttttact attttttaaa attgtagtgt atatgatagg 1235
aatttgcatt taaatatgtt catttttgca tatgttagga gtggaaacaa tctggaaaac 1295
attttttttt catccaaaaa gtattctcct tgggcatatc tgatggaaaa aaaccttgat 1355
tttattttcg tatcttttagt ctgtgttctt tctagttatt tggtagtaaat tatgtgcaat 1415
ctaaaaacac tcccacaagt atttggtttt taattataaa atcatagtat atgttctttg 1475
tagaaaactg gaaaaatata tattcaaaca gaaaaaaaaa tcaaaattcc ccataatgtt 1535
gccatctaaa aataacctct atttttagttg atatcccgtt ttcatttttg aaagccattc 1595
cttaatgcta gtttgatata cactaaaagt tttagcttaca agttcaaatt ctgccagctt 1655
ttcctgacag ctatttgcac ttttttcaga tgagtgatta ttggccattt tctttttctt 1715
tttctttatt ttatttattt atttttttga gacagagttt tgctctgttg cccaggctgg 1775
agtgcagtgg tgcaatctcg gctcactgca acctctgcct cctgggttca agtgattctc 1835
cacctcagcc tcccagtag ctgggactac ggatgcctgc caccacgcct ggctaatttt 1895
tttttgattt tttttgtaga gacggggttt caccatgttg tccaggctaa tcttgaactt 1955
gtgacctcag gtgatccacc cgctcggcc tccgaaagtg ctgggattac aggcgtgagc 2015
taccacgcc ggcttattg accattttct aaataagcac attctatctt tattctctta 2075
aaattcaaat tttctgttac tgataatcct aatactagga ttcttgctta agtatgtgaa 2135
accattaccg atttgttgtt cacatttatt ttttatgttg tgaaactgga ctaaaaggaat 2195
agagggatga ttagtataa aagtcaataa gcatttgtgt ttaactgttg agaaaaagtg 2255
aagatcagta tgattattat ggaactgttt ttaattcttg cttaaagact acaattttag 2315
tataatgaca tttgagtcta gggtagtagt tggtagattt ctagatggtc cctaattaag 2375
aagtattgtt gtatttagaa ttgtccacct aatttctttt tatataatgc caaggatttt 2435
cttgtgcttt tgggatctta tgctgtttgt aaaatgttac tgtccaatgt tggattattg 2495
ttttggtttc aggcatattg tgaatagggt atgatacatg ggtatttttc tgcaagtatt 2555
taaaccaggg gcatatgcaa aggcagttgt aatttctctt tggaaaaagc gccaaatgtt 2615
tgaaggttaa aatcaaatgc tagggttgat atttaggctt ataacaaaat aggcttgttt 2675
tcaaagcagt ttttctctag agttttaact gtttaactcac tagtttgctg ctgtttttta 2735
ctatgttaaa taacatatgg tatttggcaa atagatttat ttttcaaaat gtctcactag 2795
tttcttttta cacaatgtat ataactcaag atgtatagaa aggaaagcta cagttgagcc 2855
cttatacatg ttttaaggta gaaatatgtt ccctattgtt tgaaaactga ttgtaagaat 2915
aacctcagtt aggagatata acttgaagtg tcagtccaaa ctactgattt aaccctattt 2975
acggtaacac attaccttcc tcacctcctg tttggccctg gagaatgtag tcctttttct 3035
catttgtgtt gagaaatgaa aagtctgctg tagaatgtat ctgatgtcat tagttcttca 3095
aatggatacc attgtacata taacagtaga atttggtttg gggttgttag tgaaaaaaaa 3155
ttta

```

<210> 150
 <211> 1033
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 30..305

```

<400> 150
agaatgtcgt actctcagga ggctccacc atg ttc agg gat ttc gga cgc cga 53
                                Met Phe Arg Asp Phe Gly Arg Arg
                                1           5

ctg cag agg gat ttg aag aga gtg gtg gat gct agg ctg agg ctc agc 101
Leu Gln Arg Asp Leu Lys Arg Val Val Asp Ala Arg Leu Arg Leu Ser
    10           15           20

gag gag ctc ars sgc ggg agg atc aag ccg aag cct gtg gag gtc cag 149
Glu Glu Leu Xaa Xaa Gly Arg Ile Lys Pro Lys Pro Val Glu Val Gln

```

116

25	30	35	40	
gtg gtc acg cat	cac atg cag cgc tac	gcc gtg tgg ttc gga ggc tcc	197	
Val Val Thr His	His Met Gln Arg Tyr	Ala Val Trp Phe Gly Gly Ser		
	45	50	55	
atg ctg gcc tcg	act ccc gag ttc ttt	cag gtc tgc cac acc aag aag	245	
Met Leu Ala Ser	Thr Pro Glu Phe Phe	Gln Val Cys His Thr Lys Lys		
	60	65	70	
gac tat gaa gag	tac ggg ccc agc atc	tgc cgc cac aac ccc gtc ttt	293	
Asp Tyr Glu Glu	Tyr Gly Pro Ser Ile	Cys Arg His Asn Pro Val Phe		
	75	80	85	
gga gtc atg tcc	tagtgtctgc ctgaacgcgt	cgttcgatgg tgtcacgttg	345	
Gly Val Met Ser				
90				
gggaacaagt gtccttcaga	acccagagaa ggccgccgtt	ctgtaaataag cgacgtcggt	405	
gttgctgccc agcagcgtgc	ttgcattgcc ggtgcatgag	gcgcgggcgcg ggcccttcag	465	
taaaagccat ttatccgtgt	gccgaccgct gtctgccagc	ctcctccttc tcccgcctc	525	
ctcacccctcg ctctccctcc	tcctcctcct ccgagctgct	agctgacaaa tacaattctg	585	
aaggaatcca aatgtgactt	tgaaaattgt tagagaaaac	aacattagaa aatggcgcaa	645	
aatcgttagg tcccaggaga	gaatgtgggg gcgcaaacc	ttttcctccc agcctatatt	705	
tgtaaataaaa atgttttaaac	ttgaaataca aatcgatgtt	tatatctcct atcatattgt	765	
attttatggg atttggtaca	actggctgat actaagcacg	aatagatatt gatgttatgg	825	
agtgtctgtaa tccaaagtgt	ttaattgtga ggcatgttct	gatattgtta taggcaaaca	885	
aataaaaacag caaacttttt	tgccacatgt ttgctagaaa	atgattatac tttattggag	945	
tgacatgaag tttgaacact	aaacagtaat gtatgagaat	tactacagat acatgtatct	1005	
tttagttttt tttgtttgaa	ctttctgg		1033	

<210> 151
 <211> 1033
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 30..305

<400> 151	
agaatgtcgt actctcagga ggctccacc	atg ttc agg gat ttc gga cgc cga
	Met Phe Arg Asp Phe Gly Arg Arg
	1 5
ctg cag agg gat ttg aag aga gtg gtg gat gct agg	ctg agg ctc agc
Leu Gln Arg Asp Leu Lys Arg Val Val Asp Ala Arg	Leu Arg Leu Ser
10 15 20	
gag gag ctc agc ggc ggg agg atc aag ccg aag cct	gtg gag gtc cag
Glu Glu Leu Ser Gly Gly Arg Ile Lys Pro Lys Pro	Val Glu Val Gln
25 30 35 40	
gtg gtc acg cat cac atg cag cgc tac gcc gtg tgg ttc gga ggc tcc	197
Val Val Thr His His Met Gln Arg Tyr Ala Val Trp Phe Gly Gly Ser	
	45 50 55
atg ctg gcc tcg act ccc gag ttc ttt cag gtc tgc cac acc aag aag	245
Met Leu Ala Ser Thr Pro Glu Phe Phe Gln Val Cys His Thr Lys Lys	
	60 65 70
gac tat gaa gag tac ggg ccc agc atc tgc cgc cac aac ccc gtc ttt	293
Asp Tyr Glu Glu Tyr Gly Pro Ser Ile Cys Arg His Asn Pro Val Phe	
	75 80 85
gga gtc atg tcc tagtgtctgc ctgaacgcgt cgttcgatgg tgtcacgttg	345
Gly Val Met Ser	
90	
gggaacaagt gtccttcaga acccagagaa ggccgccgtt	ctgtaaataag cgacgtcggt
gttgctgccc agcagcgtgc ttgcattgcc ggtgcatgag	gcgcgggcgcg ggcccttcag
taaaagccat ttatccgtgt gccgaccgct gtctgccagc	ctcctccttc tcccgcctc
ctcacccctcg ctctccctcc tcctcctcct ccgagctgct	agctgacaaa tacaattctg
aaggaatcca aatgtgactt tgaaaattgt tagagaaaac	aacattagaa aatggcgcaa
aatcgttagg tcccaggaga gaatgtgggg gcgcaaacc	ttttcctccc agcctatatt
	705

117

tgtaaataaaa	atgttttaaac	ttgaaataaca	aatcgatggt	tatatttcct	atcattttgt	765
attttatggt	attttggtaca	actggctgat	actaagcacg	aatagatatt	gatgttatgg	825
agtgtctgtaa	tccaaagttt	ttaattgtga	ggcatgttct	gatatgttta	taggcaaaca	885
aataaaaacag	caaacttttt	tgccacatgt	ttgctagaaa	atgattatac	tttattggag	945
tgacatgaag	tttgaacact	aaacagtaat	gtatgagaat	tactacagat	acatgtatct	1005
tttagttttt	tttgtttgaa	ctttctgg				1033

<210> 152
 <211> 688
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 159..452

<220>
 <221> polyA_site
 <222> 678..688

<220>
 <221> polyA_signal
 <222> 670..675

<400> 152														
atacgtacac	caatgctggac	aagctcctag	aagcagcaga	gcagttggct	cagacggggg		60							
aatgtgaccc	cgaggagatc	tacaaggcag	ctcgacacct	ggaggtgcgc	atccaagact		120							
tcgtgcgcag	ggtggagcag	cggaagcttc	tcttgagac	atg tct gtt tcc ttc cac			176							
				Met Ser Val Ser Phe His										
				1	5									
aca cac acc aaa gag ttg tgg aca tgg atg gaa gac ctt cag aag gag						224								
Thr His Thr Lys Glu Leu Trp Thr Trp Met Glu Asp Leu Gln Lys Glu														
	10		15		20									
atg ttg gag gat gtc tgt gca gat tct gtg gat gca gtc cag gaa ctg						272								
Met Leu Glu Asp Val Cys Ala Asp Ser Val Asp Ala Val Gln Glu Leu														
	25		30		35									
atc aag cag ttc cag cag cag cag acc gcc act cta gat gcc aca ctc						320								
Ile Lys Gln Phe Gln Gln Gln Gln Thr Ala Thr Leu Asp Ala Thr Leu														
	40		45		50									
aat gtc atc aag gaa ggc gaa gac ctt atc cag cag ctc agg tca gcg						368								
Asn Val Ile Lys Glu Gly Glu Asp Leu Ile Gln Gln Leu Arg Ser Ala														
	55		60		65		70							
cct ccc tcc ctc ggg gag ccc agc gag gcc agg tca gca tgg gca gag						416								
Pro Pro Ser Leu Gly Glu Pro Ser Glu Ala Arg Ser Ala Trp Ala Glu														
	75		80		85									
ctt tcc agt ggg aaa tgc ctc ggg cta gac gtg aga taagtcctgt						462								
Leu Ser Ser Gly Lys Cys Leu Gly Leu Asp Val Arg														
	90		95											
cccagtcttg actcagcaaa atcttgcttt gtggtcttgt gtgagcttcg gtttctttat						522								
ctgtaagatg taaaaatgcc tgcaccatac tcaacttgaa tctagaataa aatagtaa						582								
ggaaaagctt taatgtttca aagaattatt tatcattcag gatgttattc agccatcagt						642								
aagtaacata aatacaacaa aaattaacat aaaacaaaaa aaaaaa						688								

<210> 153
 <211> 420
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 63..329

<220>

<221> polyA_site

<222> 406..420

<400> 153

```

atgctcgcca ggataccccc tcttcccatg gcgacctcag accgcgggcg tcagcgttgg      60
cg atg gtc tcc gga gac ggc ttc ctc gtt tcc agg cct gaa gcg att      107
  Met Val Ser Gly Asp Gly Phe Leu Val Ser Arg Pro Glu Ala Ile
    1         5         10        15
cat cta gga cct cgg cag gcg gtg cga cca agc gtt cgg gcc gag agc      155
His Leu Gly Pro Arg Gln Ala Val Arg Pro Ser Val Arg Ala Glu Ser
          20         25         30
cgt cga gtg gat ggt ggc ggc cgg agc cca agg gaa cca gat ggc cgg      203
Arg Arg Val Asp Gly Gly Gly Arg Ser Pro Arg Glu Pro Asp Gly Arg
          35         40         45
ggc cgg agc cgc caa gcg aga ttc tca cct tac cca atc cct gcc gtt      251
Gly Arg Ser Arg Gln Ala Arg Phe Ser Pro Tyr Pro Ile Pro Ala Val
          50         55         60
gaa ccc gat ctc cta aga agt gtg ctg caa cag cgt ttg att gca tta      299
Glu Pro Asp Leu Leu Arg Ser Val Leu Gln Gln Arg Leu Ile Ala Leu
          65         70         75
gga ggt gtt atc gca gct cga att tca gtt taaacgaaca cctttcctct      349
Gly Gly Val Ile Ala Ala Arg Ile Ser Val
          80         85
ggccctcact tagcttgtga acaggccttt ttaaaatcct ttcttggtgt agcagcaaaa      409
aaaaaaaaa a      420

```

<210> 154

<211> 640

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 181..549

<220>

<221> polyA_site

<222> 626..640

<400> 154

```

atgctcgcca gaatacccct cggttaaaggc aaggcggtt ctggctcttc cgcaggctca      60
gttatgggtt cctgtgtgct tagtcatcgg gtccgacaca ggctggactg atctggggag      120
ccgcgaaggg cctgccttca caaagggacg taacgcaagt attgcgggca gtgtttgaat      180
atg gcc ctg aac aat gtg tcc ctg tcc tcc ggt gat cag agg agc agg      228
Met Ala Leu Asn Asn Val Ser Leu Ser Ser Gly Asp Gln Arg Ser Arg
    1         5         10        15
gtg gcc tac cgc tct tcc cat ggc gac ctc aga ccg cgg gcg tca gcg      276
Val Ala Tyr Arg Ser Ser His Gly Asp Leu Arg Pro Arg Ala Ser Ala
          20         25         30
ttg gcg atg gtc tcc gga gac ggc ttc ctc gtt tcc agg cct gaa gcg      324
Leu Ala Met Val Ser Gly Asp Gly Phe Leu Val Ser Arg Pro Glu Ala
          35         40         45
att cat cta gga cct cgg cag gcg gtg cga cca agc gtt cgg gcc gag      372
Ile His Leu Gly Pro Arg Gln Ala Val Arg Pro Ser Val Arg Ala Glu
          50         55         60
agc cgt cga gtg gat ggt ggc ggc cgg agc cca agg gaa cca gat ggc      420
Ser Arg Arg Val Asp Gly Gly Gly Arg Ser Pro Arg Glu Pro Asp Gly
          65         70         75         80
cgg ggc cgg agc cgc caa gcc aga ttc tca cct tac cca atc cct gcc      468
Arg Gly Arg Ser Arg Gln Ala Arg Phe Ser Pro Tyr Pro Ile Pro Ala
          85         90         95
gtt gaa ccc gat ctc cta aga agt gtg ctg caa cag cgt ttg att gca      516
Val Glu Pro Asp Leu Leu Arg Ser Val Leu Gln Gln Arg Leu Ile Ala

```

```

      100      105      110
tta gga ggt gtt atc gca gct cga att tca gtt taaacgaaca cctttcctct 569
Leu Gly Gly Val Ile Ala Ala Arg Ile Ser Val
      115      120
ggccctcact tagcttgtga acaggccttt ttaaaatcct ttcttggtgt agcagcaaaa 629
aaaaaaaaa a 640

<210> 155
<211> 1582
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 8..523

<220>
<221> polyA_site
<222> 1567..1582

<400> 155
agtcaag atg gcg gga gca gct acc cag gct tcc ctg gag tcg gcc cca 49
      Met Ala Gly Ala Ala Thr Gln Ala Ser Leu Glu Ser Ala Pro
      1      5      10
cgg atc atg cgg ctg gtg gcc gaa tgc agc cgc tcc agg gcc cgg gca 97
Arg Ile Met Arg Leu Val Ala Glu Cys Ser Arg Ser Arg Ala Arg Ala
15      20      25      30
ggc gag ctg tgg ctg ccg cat ggg aca gtg gcc act cct gtg ttc atg 145
Gly Glu Leu Trp Leu Pro His Gly Thr Val Ala Thr Pro Val Phe Met
      35      40      45
cca gtg ggc acg cag gcc acc atg aag ggc atc acg acc gaa cag ctg 193
Pro Val Gly Thr Gln Ala Thr Met Lys Gly Ile Thr Thr Glu Gln Leu
      50      55      60
gac gct ctg ggt tgc cgc atc tgc ctg ggc aat acc tac cat ctg ggt 241
Asp Ala Leu Gly Cys Arg Ile Cys Leu Gly Asn Thr Tyr His Leu Gly
      65      70      75
cta agg ccg gga ccc gag ctg atc cag aaa gcc aac ggt ctc cac ggc 289
Leu Arg Pro Gly Pro Glu Leu Ile Gln Lys Ala Asn Gly Leu His Gly
      80      85      90
ttc atg aat tgg cct cat aat ctg cta acg gac agc ggc ggt ttc cag 337
Phe Met Asn Trp Pro His Asn Leu Leu Thr Asp Ser Gly Gly Phe Gln
95      100      105      110
atg gtg tcg ctg gtg tct ctg tcc gag gtg acg gag gag ggc gtc cgc 385
Met Val Ser Leu Val Ser Leu Ser Glu Val Thr Glu Glu Gly Val Arg
      115      120      125
ttc cgc tcc ccc tac gac ggc aat gag acc ctg ctg agc ccg gag aaa 433
Phe Arg Ser Pro Tyr Asp Gly Asn Glu Thr Leu Leu Ser Pro Glu Lys
      130      135      140
tcc gtg cag atc cag aat gcg ctg ggc tcg gac atc atc atg cag ctg 481
Ser Val Gln Ile Gln Asn Ala Leu Gly Ser Asp Ile Ile Met Gln Leu
      145      150      155
gac gac gtg gtt agc agt act gtg act ggg cca cgt gtg gag 523
Asp Asp Val Val Ser Ser Thr Val Thr Gly Pro Arg Val Glu
      160      165      170
taggccatgt acagggtcaat ccgctggctg gaccgggtgca ttgcagccca tcagcggccg 583
gacaagcaga acctotctgc cattatccag ggtgggctgg acgcagatct ccgggccacc 643
tgccttgaag agatgaccaa gcgagacgtg cctggcttcg ccatcggggg cctgagcggg 703
ggtgagagca agtcgcagtt ctggcggatg gtggcgctga gcacctctcg gctgccgaag 763
gacaagcccc gatattctgat ggggggttggg tatgttgtgg ataggggaagc cagagcccta 823
cctgtgggaa gtggattcct ggggaccccc taccctgctt ggggaggttg catttggggg 883
aaacggacac aggtctgatc tgaggagact aggaagacat ggctgtccct tgggggccat 943
tctgagggaa tatggcccag tctggggcag tgtgagtgtt gggaggggccc ctgggaagcc 1003
cctgaggttc tctgccccct cccgtcatgg ctgcaacccc agctatgccca ctgatctggt 1063

```


120

```

agtctgcgtg gctcttggat gtgacatgtt cgactgcgtc ttccccacac ggacagcgcg 1123
ctttggctct gccctgggtgc ccactgggaa cctgcagttg aggaagaagg tgtttgagaa 1183
ggacttcggc cccatagacc cggagtgcac ctgccccacg tgccaaaagc acagccgcgc 1243
cttcctgcac gcaactgctgc acagtgcaca cacggcccgcg ctgcaccacc tcacggtcca 1303
caacatcgcc taccagctgc agctcatgag cgccgtccgc accagcatcg tggagaagcg 1363
cttcccggac ttcgtgctggg acttcatggg cgccatgtac ggggatccca ccctctgtcc 1423
cacctgggcc actgacgctc tggcctctgt gggaatcaca ctgggctgac ctggcattgg 1483
gagagggagg gaggaaggaa gggagggagg ggctggaaga tactgaagga ttcctttttg 1543
aaagggtttt tttattgtaa cttaaaaaaaa aaaaaaaaaa 1582

```

<210> 156
 <211> 104
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 45..104

```

<400> 156
agagctgagc tgaagcggga cccggagccc gagcagccgc cgcc atg gca atc aaa      56
                                         Met Ala Ile Lys
                                         1
ttt ctg gaa gtc atc aag ccc ttc tgt gtc atc ctg caa cta aca tct      104
Phe Leu Glu Val Ile Lys Pro Phe Cys Val Ile Leu Gln Leu Thr Ser
5              10              15              20

```

<210> 157
 <211> 150
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 2..70

<220>
 <221> polyA_site
 <222> 135..150

<220>
 <221> polyA_signal
 <222> 115..120

```

<400> 157
t ttt ctc aac aat tcc tca ccg cag gag cct ctg aag ctc cca cca ggc      49
  Phe Leu Asn Asn Ser Ser Pro Gln Glu Pro Leu Lys Leu Pro Pro Gly
  1              5              10              15
cag ctc tcc tcc acc cac cgc taactctcag cccagtcac cctcttgag      100
Gln Leu Ser Ser Thr His Arg
                20
cttccttgct ttgaattaaa gaccactcat gcgcaaaaaa aaaaaaaaaa      150

```

<210> 158
 <211> 510
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 150..284

<220>

<221> polyA_site

<222> 495..510

<220>

<221> polyA_signal

<222> 475..480

<400> 158

```

aaccaaagca tgggctcttt gtttctcttt tgccaggcct gggtccttcc cagtgacccc      60
tcagccttca cttctctctc actcccgggg agctgtgaca ggaagaggag atcccagaac      120
tggagattga cgtggatgag ctctctggac atg gag agt gac gat gcc cgg gct      173
                               Met Glu Ser Asp Asp Ala Arg Ala
                               1           5

gcc agg gtc aag gag ctg ctg gtt gac tgt tac aaa ccc aca gag gcc      221
Ala Arg Val Lys Glu Leu Leu Val Asp Cys Tyr Lys Pro Thr Glu Ala
   10           15           20

ttc att tyt ggc ctg ctg gac aag atc cgg ggc atg cag aag ctg agc      269
Phe Ile Xaa Gly Leu Leu Asp Lys Ile Arg Gly Met Gln Lys Leu Ser
   25           30           35           40

aca ccc cag aag aag tgaggggtccc cgaccagga gaacggtggc tcccacagga      324
Thr Pro Gln Lys Lys
               45

caatcgctgc cccccaacct cgtagcaaca gcaataccgg gggaccctgc ggccaggcct      384
ggtgccatga gcagggtccc tcgtgccctt ggcccagggg tctcttcccc tgccccctca      444
gttttccact tttgggggttt ttttattgtt attaaactga tgggacttct aaaaaaaaaa      504
aaaaaa                                           510

```

<210> 159

<211> 390

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 10..336

<220>

<221> polyA_site

<222> 373..390

<220>

<221> polyA_signal

<222> 354..359

<400> 159

```

agtctgaag atg gcg gcc tca gca gcg cga ggt gct gcg gcg ctg cgt aga      51
                Met Ala Ala Ser Ala Ala Arg Gly Ala Ala Ala Leu Arg Arg
                1           5           10

agt atc aat cag ccg gtt gct ttt gtg aga aga att cct tgg act gcg      99
Ser Ile Asn Gln Pro Val Ala Phe Val Arg Arg Ile Pro Trp Thr Ala
   15           20           25           30

gcg tcg agt cag ctg aaa gaa cac ttt gca cag ttc ggc cat gtc aga      147
Ala Ser Ser Gln Leu Lys Glu His Phe Ala Gln Phe Gly His Val Arg
               35           40           45

agg tgc att tta cct ttt gac aag gag act ggc ttt cac aga ggt ttg      195
Arg Cys Ile Leu Pro Phe Asp Lys Glu Thr Gly Phe His Arg Gly Leu
               50           55           60

ggt tgg gtt cag ttt tct tca gaa gaa gga ctt cgg aat gca cta caa      243
Gly Trp Val Gln Phe Ser Ser Glu Glu Gly Leu Arg Asn Ala Leu Gln
               65           70           75

cag gaa aat cat att ata gat gga gta aag gtc cag gtt cac act aga      291
Gln Glu Asn His Ile Ile Asp Gly Val Lys Val Gln Val His Thr Arg
   80           85           90

```

122

agg cca aaa ctt ccg caa aca tct gat gat gaa aag aaa gat ttt 336
Arg Pro Lys Leu Pro Gln Thr Ser Asp Asp Glu Lys Lys Asp Phe
95 100 105
tgagactgca gcctattaat aaagttaaca taactgacaa aaaaaaaaaa aaaa 390

```
<210> 160
<211> 398
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 23..175
```

```
<220>
<221> polyA_site
<222> 383..398
```

```
<220>  
<221> polyA_signal  
<222> 365..370
```

<400>	160	
gagaaggtg	ctttagtctga	ag atg gcg gcc tca gca gcg cga ggt gct gcg
		Met Ala Ala Ser Ala Ala Arg Gly Ala Ala
	1	5 10
gcg ctg cat aag tat caa tca gcc ggt tgc ttt tgt gag aag aat tcc		100
Ala Leu His Lys Tyr Gln Ser Ala Gly Cys Phe Cys Glu Lys Asn Ser		
	15	20 25
ttg gac tgc ggc gtc gag tca gct gaa aga aca ctt tgc aca gtt cgg		148
Leu Asp Cys Gly Val Glu Ser Ala Glu Arg Thr Leu Cys Thr Val Arg		
	30	35 40
cca tgt cag aag gtg cat ttt acc ttt tgacaaggag actggctttc		195
Pro Cys Gln Lys Val His Phe Thr Phe		
	45	50
acagaggttt ggggttgggtt cagttttctt cagaagaagg acttcggaat gcactacaac		255
aggaaaatca tattatagat ggagtaaagg tccagggttca cactagaagg ccaaaacttc		315
cgcaaacatc tgatgatgaa aagaaagatt tttgagactg cagcctatta ataaagttaa		375
cataaccaaa aaaaaaaaaa aaa		398

```
<210> 161
<211> 425
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 127..423
```

```
<220>
<221> polyA_site
<222> 394..425
```

```

<400> 161
c cca gtc aga ggg ctg aac atg tcc gag ttt gtc tgt aac ctg tca gca      49
  Pro Val Arg Gly Leu Asn Met Ser Glu Phe Val Cys Asn Leu Ser Ala
    1             5             10             15
agg cct tat gtg tac gac ctc att gcc gtg tcc aat cat tat gga gcc      97
Arg Pro Tyr Val Tyr Asp Leu Ile Ala Val Ser Asn His Tyr Gly Ala
      20             25             30
atg ggg gtt ggc cac tac taaagcagct tatgtgctat tttaaccaacg      145
Met Gly Val Gly His Tyr
      35

```

123

tcgagatgat	gaattttata	agacaccttc	acttagcagt	tctggttcct	ctgatggagg	205
gacacgacca	agcagctctc	agcagggtt	tggggatgat	gaggcttgca	gcatggacac	265
caactaatgc	tgactccacg	atcctgccac	cctgtagcgc	cagtgtaatc	ccccaggaga	325
acatctttga	cactctgcag	actgctagt	ttctgtctaa	aaaccagaca	aggaaatacc	385
cttcttttat	gagcagaagg	aawcaaaaaa	aaaaaaaaaa			425

<210> 162
 <211> 438
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 3..41

<220>
 <221> polyA_site
 <222> 419..438

<400> 162		
ag atg tta ggc ttt gcc aaa aac tgg atc ttt agt atc act tgagcactca		51
Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr		
1 5 10		
ttggacacaa gaggtcaact gttttgtcat ttgtgtagta agaacaaaat aaagtgtgtg		111
gcattaaaga aataacactg tctggccggg atcgggtggct cacgcctgta atcccagcac		171
tttgggaggc cgaggcgggc agatcacgag gtcaggagtt cgagaccagc ctgaccaaca		231
tgatgaaacc ctgtctctac taaaaataca aaaattagcc aggcattggtg gtgcacacct		291
gtagtcccag ctgctcagga ggctgaggca ggagaatcgc tcgaaccggg gaggtggagg		351
ttgcagttag ctgagatggc gccactgcat tccaggcctg gggggtgaca gagcaagact		411
cagtctcaga aaaaaaaaaa aaaaaaag		438

<210> 163
 <211> 436
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 3..41

<220>
 <221> polyA_site
 <222> 419..436

<400> 163		
ag atg tta ggc ttt gcc aaa aac tgg atc ttt agt atc act tgagcactca		51
Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr		
1 5 10		
ttggacacaa gaggtcaact gttttgtcat ttgtgtagta agaacaaaat aaagtgtgtg		111
gcattaaaga aataacactg tctggccggg atcgggtggct cacgcctgta atcccagcac		171
tttgggaggc cgaggcgggc agatcacgag gtcaggagtt cgagaccagc ctgaccaaca		231
tgatgaaacc ctgtctctac taaaaataca aaaattagcc aggcattggtg gtgcacacct		291
gtagtcccag ctgctcagga ggctgaggca ggagaatcgc tcgaaccggg gaggtggagg		351
ttgcagttag ctgagatggc gccactgcat tccaggcctg gggggtgaca gagcaagact		411
cagtctcaga aaaaaaaaaa aaaaaa		436

<210> 166
 <211> 436
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS

<222> 3..41

<220>

<221> polyA_site

<222> 421..436

<400> 166

```

ag atg tta ggc ttt gcc aaa aac tgg atc ttt agt atc act tgagcactca      51
  Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr
      1             5             10
ttggacacaa gaggtcaact gttttgtcat ttgtgtagta agaacaaaaat aaagtgtgtg      111
gcattaaaga aataacactg tctggccggg atcgggtggct cagcctgta atcccagcac      171
tttgggaggg cgaggcgggc agatcacgag gtcaggagtt cgagaccagc ctgaccaaca      231
tgatgaaacc ctgtctctac taaaaatata aaaattagcc aggcattggtg gtgcacacct      291
gtagtcccag ctgctcagga ggctgaggca ggagaatcgc tcgaaccccg gaggtggagg      351
ttgcagttag ctgagatggc gccactgcat tccaggcctg gggggtgaca gagcaagact      411
cagtctcaga aaaaaaaaaa aaaaaa

```

<210> 167

<211> 436

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 3..41

<220>

<221> polyA_site

<222> 419..436

<400> 167

```

ag atg tta ggc ttt gcc aaa aac tgg atc ttt agt atc act tgagcactca      51
  Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr
      1             5             10
ttggacacaa gaggtcaact gttttgtcat ttgtgtagta agaacaaaaat taagtgtgtg      111
gcattaaaga aataacactg tctggccggg atcgggtggct cagcctgta atcccagcac      171
tttgggaggg cgaggcgggc agatcacgag gtcaggagtt cgagaccagc ctgaccaaca      231
tgatgaaacc ctgtctctac taaaaatata aaaattagcc aggcattggtg gtgcacacct      291
gtagtcccag ctgctcagga ggctgaggca ggagaatcgc tcgaaccccg gaggtggagg      351
ttgcagttag ctgagatggc gccactgcat tccaggcctg gggggtgaca gagcaagact      411
cagtctcaga aaaaaaaaaa aaaaaa

```

<210> 168

<211> 394

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 10..336

<220>

<221> polyA_site

<222> 373..394

<220>

<221> polyA_signal

<222> 354..359

<400> 168

```

agtctgaag atg gcg gcc tca gca gcg cga ggt gct gcg gcg ctg cgt aga      51
  Met Ala Ala Ser Ala Ala Arg Gly Ala Ala Ala Leu Arg Arg

```

125

	1		5		10												
agt	atc	aat	cag	ccg	gtt	gct	ttt	gtg	aga	aga	atw	cct	tgg	act	gcg		99
Ser	Ile	Asn	Gln	Pro	Val	Ala	Phe	Val	Arg	Arg	Ile	Pro	Trp	Thr	Ala		
15					20					25					30		
gcg	tcg	agt	cag	ctg	aaa	gaa	cac	ttt	gca	cag	ttc	ggc	cat	gtc	aga		147
Ala	Ser	Ser	Gln	Leu	Lys	Glu	His	Phe	Ala	Gln	Phe	Gly	His	Val	Arg		
			35					40					45				
agg	tgc	att	tta	cct	ttt	gac	aag	gag	act	ggc	ttt	cac	aga	ggg	ttg		195
Arg	Cys	Ile	Leu	Pro	Phe	Asp	Lys	Glu	Thr	Gly	Phe	His	Arg	Gly	Leu		
			50					55					60				
ggg	tgg	gtt	cag	ttt	tct	tca	gaa	gaa	gga	ctt	cgg	aat	gca	cta	caa		243
Gly	Trp	Val	Gln	Phe	Ser	Ser	Glu	Glu	Gly	Leu	Arg	Asn	Ala	Leu	Gln		
			65				70						75				
cag	gaa	aat	cat	att	ata	gat	gga	gta	aag	gtc	cag	gtt	cac	act	aga		291
Gln	Glu	Asn	His	Ile	Ile	Asp	Gly	Val	Lys	Val	Gln	Val	His	Thr	Arg		
			80				85				90						
agg	cca	aaa	ctt	ccg	caa	aca	tct	gat	gat	gaa	aag	aaa	gat	ttt			336
Arg	Pro	Lys	Leu	Pro	Gln	Thr	Ser	Asp	Asp	Glu	Lys	Lys	Asp	Phe			
95					100					105							
tgagactgca	gcctattaat	aaagttaaca	taactgaaaa	aaaaaaaaaa	aaggtcaa												394

<210> 169

<211> 772

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 595..771

<400> 169

atataggcgt	ggctgaatac	gggcctcctg	cgtacgtgcg	tggtgtacgt	acgtgcgtga		60
ttacgcgcac	acgtacgttc	ctcatgaaag	ggacgacggg	agctgcatga	aagccgaagt		120
tatggaccgc	tagcatctgt	cactggccac	cggtttcctg	gagtaagcgg	cagctacctt		180
acagccctga	cacgagccgg	gtgctctctc	ttctcaccgc	ggcccacgtc	tcctcgctgg		240
ctccggtggc	ctcgctgggt	cgcgaggagg	cggaggactg	tactctgagg	ccaaaagcca		300
gagtcggccc	tgaacgcccc	cgactctcag	ggtccagagg	ccgtgagacc	ggccgcgggt		360
gaaaggtaaa	gaaaccaagt	ggaagagtgt	ttcctctctt	ggccgtaaaag	caggtactct		420
ctgcagcacc	agctgtcccc	gccctactcc	ggaccgcccc	aaagactcca	tgggatggac		480
ctgagtcagc	cgaatcctag	ccccttccct	tgggcctgct	gtggtgctcg	acatcagtga		540
cagacggaag	cagcagacca	tcaaggctac	gggaggcccc	gggcgcttgc	gaag atg		597

Met

1

aag	ttt	ggc	tgc	ctc	tcc	ttc	cgg	cag	cct	tat	gct	ggc	ttt	gtc	tta		645
Lys	Phe	Gly	Cys	Leu	Ser	Phe	Arg	Gln	Pro	Tyr	Ala	Gly	Phe	Val	Leu		
			5					10					15				
aat	gga	atc	aag	act	gtg	gag	acg	cgc	tgg	cgt	cct	ctg	ytg	agc	agc		693
Asn	Gly	Ile	Lys	Thr	Val	Glu	Thr	Arg	Trp	Arg	Pro	Leu	Leu	Ser	Ser		
			20				25					30					
cag	cgg	aac	tgt	acc	atc	gcc	gtc	cac	att	gct	cac	agg	gac	tgg	gaa		741
Gln	Arg	Asn	Cys	Thr	Ile	Ala	Val	His	Ile	Ala	His	Arg	Asp	Trp	Glu		
			35				40					45					
ggc	gat	gcc	tgt	cgg	gag	ctg	ctg	gtg	gag	a							772
Gly	Asp	Ala	Cys	Arg	Glu	Leu	Leu	Val	Glu								
50					55												

<210> 170

<211> 128

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -34..-1

<220>

<221> UNSURE

<222> 7

<223> Xaa = Ile,Thr

<400> 170

Met	Ser	Leu	Arg	Leu	Asp	Thr	Thr	Pro	Ser	Cys	Asn	Ser	Ala	Arg	Pro
				-30					-25					-20	
Leu	His	Ala	Leu	Gln	Val	Leu	Leu	Leu	Leu	Ser	Leu	Leu	Leu	Thr	Ala
			-15					-10					-5		
Leu	Ala	Ser	Ser	Thr	Lys	Gly	Gln	Xaa	Lys	Arg	Asn	Leu	Ala	Lys	Gly
		1				5					10				
Lys	Asp	Glu	Ser	Leu	Asp	Ser	Asp	Leu	Tyr	Ala	Glu	Leu	Arg	Cys	Met
15					20					25				30	
Cys	Ile	Lys	Thr	Thr	Ser	Gly	Ile	His	Pro	Lys	Asn	Ile	Gln	Ser	Leu
				35					40					45	
Glu	Val	Ile	Gly	Lys	Gly	Thr	His	Cys	Asn	Gln	Val	Glu	Val	Ile	Ala
			50					55					60		
Thr	Leu	Lys	Asp	Gly	Arg	Lys	Ile	Cys	Leu	Asp	Pro	Asp	Ala	Pro	Arg
			65				70					75			
Ile	Lys	Lys	Ile	Val	Gln	Lys	Lys	Leu	Ala	Gly	Asp	Glu	Ser	Ala	Asp
	80					85					90				

<210> 171

<211> 160

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -18..-1

<400> 171

Met	Lys	Ala	Ala	Val	Leu	Thr	Leu	Ala	Val	Leu	Phe	Leu	Thr	Gly	Ser
			-15					-10					-5		
Gln	Ala	Arg	His	Phe	Trp	Gln	Gln	Asp	Glu	Pro	Pro	Gln	Ser	Pro	Trp
		1				5					10				
Asp	Arg	Val	Lys	Asp	Leu	Ala	Thr	Val	Tyr	Val	Asp	Val	Leu	Lys	Asp
15					20				25					30	
Ser	Gly	Arg	Asp	Tyr	Val	Ser	Gln	Phe	Glu	Gly	Ser	Ala	Leu	Gly	Lys
			35						40					45	
Gln	Leu	Asn	Leu	Lys	Leu	Leu	Asp	Asn	Trp	Asp	Ser	Val	Thr	Ser	Ala
			50					55					60		
Phe	Ser	Asn	Leu	Arg	Glu	Gln	Leu	Gly	Pro	Val	Thr	Gln	Glu	Phe	Trp
			65				70					75			
Asp	Asn	Leu	Glu	Lys	Glu	Thr	Glu	Gly	Leu	Arg	Gln	Glu	Met	Ser	Lys
80						85					90				
Asp	Leu	Glu	Glu	Val	Lys	Ala	Lys	Val	Gln	Pro	Tyr	Leu	Asp	Asp	Phe
95					100					105					110
Gln	Lys	Lys	Trp	Gln	Glu	Glu	Met	Glu	Leu	Tyr	Arg	Gln	Lys	Ala	Ala
				115					120					125	
Phe	Leu	Thr	Ile	Leu	Thr	Ser	Leu	Gly	Pro	Asn	Gly	Asn	Lys	Ala	Phe
			130					135					140		

<210> 172

<211> 80

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -26..-1

<220>

<221> UNSURE

<222> 52

<223> Xaa = Leu, Pro

<400> 172

```

Met Gly Arg Ala Met Val Ala Arg Leu Gly Leu Gly Leu Leu Leu Leu
  -25                      -20                      -15
Ala Leu Leu Leu Pro Thr Gln Ile Tyr Ser Ser Glu Thr Thr Thr Gly
-10                      -5                      1                      5
Thr Ser Ser Asn Ser Ser Gln Ser Thr Ser Asn Ser Gly Leu Ala Pro
                      10                      15                      20
Asn Pro Thr Asn Ala Thr Thr Lys Val Ala Gly Gly Ala Leu Gln Ser
                      25                      30                      35
Thr Ala Ser Leu Phe Val Val Ser Leu Ser Leu Leu His Xaa Tyr Ser
  40                      45                      50

```

<210> 173

<211> 101

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -27..-1

<223> Von Heijne matrix
 score 11.8486806626293
 seq LFLGLLLLPLVVA/FA

<400> 173

```

Met Ser Ser Ala Ala Gly Phe Cys Ala Ser Arg Pro Gly Leu Leu Phe
  -25                      -20                      -15
Leu Gly Leu Leu Leu Leu Pro Leu Val Val Ala Phe Ala Ser Ala Glu
-10                      -5                      1                      5
Ala Glu Glu Asp Gly Asp Leu Gln Cys Leu Cys Val Lys Thr Thr Ser
                      10                      15                      20
Gln Val Arg Pro Arg His Ile Thr Ser Leu Glu Val Ile Lys Ala Gly
                      25                      30                      35
Pro His Cys Pro Thr Ala Gln Leu Ile Ala Thr Leu Lys Asn Gly Arg
                      40                      45                      50
Lys Ile Cys Leu Asp Leu Gln Ala Pro Leu Tyr Lys Lys Ile Ile Lys
  55                      60                      65
Lys Leu Leu Glu Ser
  70

```

<210> 174

<211> 101

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -27..-1

<223> Von Heijne matrix
 score 11.8486806626293
 seq LFLGLLLLPLVVA/FA

<400> 174

```

Met Ser Ser Ala Ala Gly Phe Cys Ala Ser Arg Pro Gly Leu Leu Phe
  -25                      -20                      -15
Leu Gly Leu Leu Leu Leu Pro Leu Val Val Ala Phe Ala Ser Ala Glu

```


128															
-10				-5				1				5			
Ala	Glu	Glu	Asp	Gly	Asp	Leu	Gln	Cys	Leu	Cys	Val	Lys	Thr	Thr	Ser
10				15				20							
Gln	Val	Arg	Pro	Arg	His	Ile	Thr	Ser	Leu	Glu	Val	Ile	Lys	Ala	Gly
25				30				35							
Pro	His	Cys	Pro	Thr	Ala	Gln	Leu	Ile	Ala	Thr	Leu	Lys	Asn	Gly	Arg
40				45				50							
Lys	Ile	Cys	Leu	Asp	Leu	Gln	Ala	Pro	Leu	Tyr	Lys	Lys	Ile	Ile	Lys
55				60				65							
Lys	Leu	Leu	Glu	Ser											
70															

```
<210> 175
<211> 86
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -18..-1
```

```

<400> 175
Met Ile Pro Ala Val Val Leu Leu Leu Leu Leu Leu Val Glu Gln Ala
      -15      -10      -5
Ala Ala Leu Gly Glu Pro Gln Leu Cys Tyr Ile Leu Asp Ala Ile Leu
      1          5          10
Phe Leu Tyr Gly Ile Val Leu Thr Leu Leu Tyr Cys Arg Leu Lys Ile
15      20      25      30
Gln Val Arg Lys Ala Ala Ile Thr Ser Tyr Glu Lys Ser Asp Gly Val
      35      40      45
Tyr Thr Gly Leu Ser Thr Arg Asn Gln Glu Thr Tyr Glu Thr Leu Lys
      50      55      60
His Glu Lys Pro Pro Gln
      65

```

```
<210> 176
<211> 92
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -24..-1
```

```
<220>
<221> UNSURE
<222> 52
<223> Xaa = Arg,Thr
```

<400> 176															
Met	Val	Met	Gly	Leu	Gly	Val	Leu	Leu	Leu	Val	Phe	Val	Leu	Gly	Leu
				-20					-15					-10	
Gly	Leu	Thr	Pro	Pro	Thr	Leu	Ala	Gln	Asp	Asn	Ser	Arg	Tyr	Thr	His
			-5					1				5			
Phe	Leu	Thr	Gln	His	Tyr	Asp	Ala	Lys	Pro	Gln	Gly	Arg	Asp	Asp	Arg
	10					15					20				
Tyr	Cys	Glu	Ser	Ile	Met	Arg	Arg	Arg	Gly	Leu	Thr	Ser	Pro	Cys	Lys
25				30						35					40
Asp	Ile	Asn	Thr	Phe	Ile	His	Gly	Asn	Lys	Arg	Xaa	Ser	Arg	Pro	Ser
				45					50					55	
Val	Lys	Thr	Arg	Met	Glu	Thr	Leu	Thr	Glu	Lys	Thr				
			60					65							

<210> 177
 <211> 111
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -16..-1

<400> 177
 Met Leu Leu Ile Leu Leu Ser Val Ala Leu Leu Ala Phe Ser Ser Ala
 -15 -10 -5
 Gln Asp Leu Asp Glu Asp Val Ser Gln Glu Asp Val Pro Leu Val Ile
 1 5 10 15
 Ser Asp Gly Gly Asp Ser Glu Gln Phe Ile Asp Glu Glu Arg Gln Gly
 20 25 30
 Pro Pro Leu Gly Gly Gln Gln Ser Gln Pro Ser Ala Gly Asp Gly Asn
 35 40 45
 Gln Asn Asp Gly Pro Gln Gln Gly Pro Pro Gln Gln Gly Gly Gln Gln
 50 55 60
 Gln Gln Gly Pro Pro Pro Pro Gln Gly Lys Pro Gln Gly Pro Pro Pro
 65 70 75 80
 Gln Gly Gly Arg Pro Gln Gly Pro Pro Gln Gly Gln Ser Pro Gln
 85 90 95

<210> 178
 <211> 106
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -16..-1

<400> 178
 Met Leu Leu Ile Leu Leu Ser Val Ala Leu Leu Ala Phe Ser Ser Ala
 -15 -10 -5
 Gln Asp Leu Asp Glu Asp Val Ser Gln Glu Asp Val Pro Leu Val Ile
 1 5 10 15
 Ser Asp Gly Gly Asp Ser Glu Gln Phe Ile Asp Glu Glu Arg Gln Gly
 20 25 30
 Pro Pro Leu Gly Gly Gln Gln Ser Gln Pro Ser Ala Gly Asp Gly Asn
 35 40 45
 Gln Asp Asp Gly Pro Gln Gln Gly Pro Pro Gln Gln Gly Gly Gln Gln
 50 55 60
 Gln Gln Gly Pro Pro Pro Pro Gln Gly Lys Pro Gln Gly Pro Pro Gln
 65 70 75 80
 Gln Gly Gly Gln Ser Cys Cys Cys Asp Lys
 85 90

<210> 179
 <211> 141
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -24..-1

<400> 179
 Met Pro Ser Ser Val Ser Trp Gly Ile Leu Leu Leu Ala Gly Leu Cys
 -20 -15 -10
 Cys Leu Val Pro Val Ser Leu Ala Glu Asp Pro Gln Gly Asp Ala Ala

130

			-5					1			5					
Gln	Lys	Thr	Asp	Thr	Ser	His	His	Asp	Gln	Asp	His	Pro	Thr	Phe	Asn	
	10					15					20					
Lys	Ile	Thr	Pro	Asn	Leu	Ala	Glu	Phe	Ala	Phe	Ser	Leu	Tyr	Arg	Gln	
25					30					35					40	
Leu	Ala	His	Gln	Ser	Asn	Ser	Thr	Asn	Ile	Phe	Phe	Ser	Pro	Val	Ser	
			45						50					55		
Ile	Ala	Thr	Ala	Phe	Ala	Met	Leu	Ser	Leu	Gly	Thr	Lys	Ala	Asp	Thr	
			60						65				70			
His	Asp	Glu	Ile	Leu	Glu	Gly	Leu	Asn	Phe	Asn	Leu	Thr	Glu	Ile	Pro	
		75					80					85				
Glu	Ala	Gln	Ile	His	Glu	Gly	Phe	Gln	Glu	Leu	Leu	Arg	Thr	Leu	Asn	
	90					95					100					
Gln	Pro	Asp	Ser	Gln	Leu	Gln	Leu	Thr	Thr	Gly	Lys	Asn				
105					110					115						

<210> 180

<211> 124

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -30..-1

<400> 180

Met	Pro	Ala	Cys	Arg	Leu	Gly	Pro	Leu	Ala	Ala	Ala	Leu	Leu	Leu	Ser	
-30					-25					-20					-15	
Leu	Leu	Leu	Phe	Gly	Phe	Thr	Leu	Val	Ser	Gly	Thr	Gly	Ala	Glu	Lys	
			-10						-5					1		
Thr	Gly	Val	Cys	Pro	Glu	Leu	Gln	Ala	Asp	Gln	Asn	Cys	Thr	Gln	Glu	
		5				10					15					
Cys	Val	Ser	Asp	Ser	Glu	Cys	Ala	Asp	Asn	Leu	Lys	Cys	Cys	Ser	Ala	
	20					25					30					
Gly	Cys	Ala	Thr	Phe	Cys	Ser	Leu	Pro	Asn	Asp	Lys	Glu	Gly	Ser	Cys	
35					40					45					50	
Pro	Gln	Val	Asn	Ile	Asn	Phe	Pro	Gln	Leu	Gly	Leu	Cys	Arg	Asp	Gln	
			55						60					65		
Cys	Gln	Val	Asp	Ser	Gln	Cys	Pro	Gly	Gln	Met	Lys	Cys	Cys	Arg	Asn	
			70					75					80			
Gly	Cys	Gly	Lys	Val	Ser	Cys	Val	Thr	Pro	Asn	Phe					
		85					90									

<210> 181

<211> 124

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -30..-1

<400> 181

Met	Pro	Ala	Cys	Arg	Leu	Gly	Pro	Leu	Ala	Ala	Ala	Leu	Leu	Leu	Ser	
-30					-25					-20					-15	
Leu	Leu	Leu	Phe	Gly	Phe	Thr	Leu	Val	Ser	Gly	Thr	Gly	Ala	Glu	Lys	
			-10						-5					1		
Thr	Gly	Val	Cys	Pro	Glu	Leu	Gln	Ala	Asp	Gln	Asn	Cys	Thr	Gln	Glu	
		5				10					15					
Cys	Val	Ser	Asp	Ser	Glu	Cys	Ala	Asp	Asn	Ile	Lys	Cys	Cys	Ser	Ala	
	20					25					30					
Gly	Cys	Ala	Thr	Phe	Cys	Ser	Leu	Pro	Asn	Asp	Lys	Glu	Gly	Ser	Cys	
35					40					45					50	

131

Pro Gln Val Asn Ile Asn Phe Pro Gln Leu Gly Leu Cys Arg Asp Gln
 55 60 65
 Cys Gln Val Asp Ser Gln Cys Pro Gly Gln Met Lys Cys Cys Arg Asn
 70 75 80
 Gly Cys Gly Lys Val Ser Cys Val Thr Pro Asn Phe
 85 90

<210> 182

<211> 69

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -17..-1

<223> Von Heijne matrix
 score 9.5148977432302
 seq FVFLLFVISLAAA/AH

<400> 182

Met Phe Pro Gly Phe Val Phe Leu Leu Phe Val Ile Ser Leu Ala Ala
 -15 -10 -5
 Ala Ala His Leu Trp Val Leu Ala Ala Phe Met Gly Arg Ile Thr Val
 1 5 10 15
 Lys Val Cys Ser Phe Thr Pro Glu Ala Ser Lys Thr Val Ser Pro Pro
 20 25 30
 Glu Gly Ala Asn Asn Ser Arg Arg Thr Ala Phe Lys Ser Cys Asn Thr
 35 40 45
 His His Lys Gly Leu
 50

<210> 183

<211> 77

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -33..-1

<220>

<221> UNSURE

<222> 6

<223> Xaa = Lys, Met

<400> 183

Met Gly Pro Val Lys Gln Leu Lys Arg Met Phe Glu Pro Thr Arg Leu
 -30 -25 -20
 Ile Ala Thr Ile Met Val Leu Leu Ser Phe Ala Leu Thr Leu Cys Ser
 -15 -10 -5
 Ala Phe Trp Trp His Asn Xaa Gly Leu Ala Leu Ile Phe Cys Ile Leu
 1 5 10 15
 Gln Ser Leu Ala Leu Thr Trp Tyr Ser Leu Ser Phe Ile Pro Phe Ala
 20 25 30
 Arg Asp Ala Val Lys Lys Cys Phe Ala Val Cys Leu Ala
 35 40

<210> 184

<211> 146

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL
 <222> -26..-1

<400> 184

```

Met Thr Met Arg Ser Leu Leu Arg Thr Pro Phe Leu Cys Gly Leu Leu
  -25                      -20                      -15
Trp Ala Phe Cys Ala Pro Gly Ala Arg Ala Glu Glu Pro Ala Ala Ser
-10                      -5                      1                      5
Phe Ser Gln Pro Gly Ser Met Gly Leu Asp Lys Asn Thr Val His Asp
                      10                      15                      20
Gln Glu His Ile Met Glu His Leu Glu Gly Val Ile Asn Lys Pro Glu
                      25                      30                      35
Ala Glu Met Ser Pro Gln Glu Leu Gln Leu His Tyr Phe Lys Met His
  40                      45                      50
Asp Tyr Asp Gly Asn Asn Leu Leu Asp Gly Leu Glu Leu Ser Thr Ala
55                      60                      65                      70
Ile Thr His Val His Lys Glu Glu Gly Ser Glu Gln Ala Pro Leu Met
                      75                      80                      85
Ser Glu Asp Glu Leu Ile Asn Ile Ile Asp Gly Val Leu Arg Asp Asp
                      90                      95                      100
Asp Lys Asn Asn Asp Gly Tyr Ile Asp Tyr Ala Glu Phe Ala Lys Ser
                      105                      110                      115
Leu Gln
  120

```

<210> 185
 <211> 62
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -46..-1
 <223> Von Heijne matrix
 score 8.64329745298384
 seq AVLILLILFAIVFG/LL

<400> 185

```

Met Asp Gln Leu Val Phe Lys Glu Thr Ile Trp Asn Asp Ala Phe Trp
  -45                      -40                      -35
Gln Asn Pro Trp Asp Gln Gly Gly Leu Ala Val Ile Ile Leu Phe Ile
-30                      -25                      -20                      -15
Thr Ala Val Leu Leu Ile Leu Phe Ala Ile Val Phe Gly Leu Leu
                      -10                      -5                      1
Thr Ser Thr Glu Asn Thr Gln Cys Glu Ala Gly Glu Glu Glu
                      5                      10                      15

```

<210> 186
 <211> 62
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -46..-1
 <223> Von Heijne matrix
 score 8.64329745298384
 seq AVLILLILFAIVFG/LL

<400> 186

```

Met Asp Gln Leu Val Phe Lys Glu Thr Ile Trp Asn Asp Ala Phe Trp
  -45                      -40                      -35
Gln Asn Pro Trp Asp Gln Gly Gly Leu Ala Val Ile Ile Leu Phe Ile

```

133

-30 -25 -20 -15
 Thr Ala Val Leu Leu Leu Ile Leu Phe Ala Ile Val Phe Gly Leu Leu
 -10 -5 1
 Thr Ser Thr Glu Asn Thr Gln Cys Glu Ala Gly Glu Glu Glu
 5 10 15

<210> 187
 <211> 162
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

<400> 187
 Met Arg Ile Ala Asn Arg Thr Arg Phe Ser Ser Pro Phe Leu Ala Arg
 -45 -40 -35
 Gly Ala Gly Trp Thr His Gly Arg Gly Met Met Val Val Gly Thr Gly
 -30 -25 -20
 Thr Ser Leu Ala Leu Ser Ser Leu Arg Ser Leu Leu Leu Phe Ala Gly
 -15 -10 -5 1
 Met Gln Met Tyr Ser Arg Gln Leu Ala Ser Thr Glu Trp Leu Thr Ile
 5 10 15
 Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe Val Phe Ser Leu Thr Ala
 20 25 30
 Phe Asn Asn Leu Glu Asn Leu Val Phe Gly Lys Gly Phe Gln Ala Lys
 35 40 45
 Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu Leu Ala Leu Phe Ala Ser
 50 55 60 65
 Gly Leu Ile His Arg Val Cys Val Thr Thr Cys Phe Ile Phe Ser Met
 70 75 80
 Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser Ser Thr Leu Tyr Gln Ala
 85 90 95
 Ala Ala Pro Val Leu Thr Pro Ala Lys Val Thr Gly Lys Ser Lys Lys
 100 105 110
 Arg Asn
 115

<210> 188
 <211> 90
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -35..-1

<400> 188
 Met Glu Thr Gly Ala Ser Ala Ser Ile Pro Glu Leu Ile Cys Glu Ala
 -35 -30 -25 -20
 Met Arg Arg Ile Trp Ser Leu Gly Leu Gly Leu Val Thr Leu Thr Ala
 -15 -10 -5
 Ser Trp Ala Ala Leu Phe His Asp Gly Phe Ala Val Leu Gly Gly Asn
 1 5 10
 Ile Val Ser Asp Leu Ser Thr Val Arg Phe Val Ala Gln Gln Gln His
 15 20 25
 Phe Gln Leu Leu Asp Val Trp Thr Arg Asn Phe Arg Lys Pro Leu Gly
 30 35 40 45
 Ser Met Cys Phe Val Phe Leu Leu Leu Pro
 50 55

<210> 189

<211> 146
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -25..-1

<220>
 <221> UNSURE
 <222> 60
 <223> Xaa = Lys,Arg

<400> 189
 Met Arg Leu Leu Gln Leu Leu Phe Arg Ala Ser Pro Ala Thr Leu Leu
 -25 -20 -15 -10
 Leu Val Leu Cys Leu Gln Leu Gly Ala Asn Lys Ala Gln Asp Asn Thr
 -5 1 5
 Arg Lys Ile Ile Ile Lys Asn Phe Asp Ile Pro Lys Ser Val Arg Pro
 10 15 20
 Asn Asp Glu Val Thr Ala Val Leu Ala Val Gln Thr Glu Leu Lys Glu
 25 30 35
 Cys Met Val Val Lys Thr Tyr Leu Ile Ser Ser Ile Pro Leu Gln Gly
 40 45 50 55
 Ala Phe Asn Tyr Xaa Tyr Thr Ala Cys Leu Cys Asp Asp Asn Pro Lys
 60 65 70
 Thr Phe Tyr Trp Asp Phe Tyr Thr Asn Arg Thr Val Gln Ile Ala Ala
 75 80 85
 Val Val Asp Val Ile Arg Glu Leu Gly Ile Cys Pro Asp Asp Ala Ala
 90 95 100
 Val Ile Pro Ile Lys Asn Asn Arg Phe Tyr Thr Ile Glu Ile Leu Lys
 105 110 115
 Val Glu
 120

<210> 190
 <211> 120
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -32..-1
 <223> Von Heijne matrix
 score 7.64815044221151
 seq VTLGIGFFALASA/LW

<400> 190
 Met Ala Ser Pro Lys Gly Phe Phe Asn Tyr Leu Thr Tyr Phe Leu Ala
 -30 -25 -20
 Ala Gly Ala Val Thr Leu Gly Ile Gly Phe Phe Ala Leu Ala Ser Ala
 -15 -10 -5
 Leu Trp Phe Leu Ile Cys Lys Arg Arg Glu Ile Phe Gln Asn Ser Lys
 1 5 10 15
 Phe Lys Ala Ile Asp Glu Arg Cys Arg Gln Arg Pro Ser Met Ala Lys
 20 25 30
 Ile Lys Ser His Ser Gln Cys Val Phe Ile Ser Arg Asn Phe His Thr
 35 40 45
 Gly Arg Phe Gln Leu Gln Gln Leu Lys Ile Ile Leu Lys Met Asn Pro
 50 55 60
 Asn Leu Gln Gln Lys Ile Ser Phe Val Ile Pro Gln Arg Pro Ala Pro
 65 70 75 80
 Gln Gln Ile Ala Ala Val Leu His

85

<210> 191
 <211> 78
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19..-1

<400> 191
 Met Lys Phe Phe Val Phe Ala Leu Val Leu Ala Leu Met Ile Ser Met
 -15 -10 -5
 Ile Ser Ala Asp Ser His Glu Lys Arg His His Gly Tyr Arg Arg Lys
 1 5 10
 Phe His Glu Lys His His Ser Tyr His Ile Thr Leu Leu Pro Leu Phe
 15 20 25
 Glu Glu Ser Ser Lys Ser Asn Ala Asn Glu Lys His Tyr Asn Leu Leu
 30 35 40 45
 Tyr Thr Leu Cys Phe Arg Ile Leu Ala Phe Ser Ile Val Thr
 50 55

<210> 192
 <211> 87
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -20..-1
 <223> Von Heijne matrix
 score 7.42708462116258
 seq LLVFLAGFPVLDA/ND

<400> 192
 Met Gln Lys Val Thr Leu Gly Leu Leu Val Phe Leu Ala Gly Phe Pro
 -20 -15 -10 -5
 Val Leu Asp Ala Asn Asp Leu Glu Asp Lys Asn Ser Pro Phe Tyr Tyr
 1 5 10
 Asp Trp His Ser Leu Gln Val Gly Gly Leu Ile Cys Ala Gly Val Leu
 15 20 25
 Cys Ala Met Gly Ile Ile Ile Val Met Ser Ala Lys Cys Lys Cys Lys
 30 35 40
 Phe Gly Gln Lys Ser Gly His His Pro Gly Glu Thr Pro Pro Leu Ile
 45 50 55 60
 Thr Pro Gly Ser Ala Gln Ser
 65

<210> 193
 <211> 105
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19..-1
 <223> Von Heijne matrix
 score 7.20796835452081
 seq VSIMLLLVTVSDC/AV

<220>
 <221> unsure

<222> 37

<223> Xaa = * ,Glu

<400> 193

```

Met Arg Gly Ala Thr Arg Val Ser Ile Met Leu Leu Leu Val Thr Val
              -15              -10              -5
Ser Asp Cys Ala Val Ile Thr Gly Ala Cys Glu Arg Asp Val Gln Cys
              1              5              10
Gly Ala Gly Thr Cys Cys Ala Ile Ser Leu Trp Leu Arg Gly Leu Arg
      15              20              25
Met Cys Thr Pro Leu Gly Arg Xaa Gly Glu Glu Cys His Pro Gly Ser
30              35              40              45
His Lys Ile Pro Phe Phe Arg Lys Arg Lys His His Thr Cys Pro Cys
              50              55              60
Leu Pro Asn Leu Leu Cys Ser Arg Phe Pro Asp Gly Arg Tyr Arg Cys
              65              70              75
Ser Met Asp Leu Lys Asn Ile Asn Phe
      80              85

```

<210> 194

<211> 105

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -19...-1

```

<223> Von Heijne matrix
      score 7.20796835452081
      seq VSIMLLLVTVSDC/AV

```

<400> 194

```

Met Arg Gly Ala Thr Arg Val Ser Ile Met Leu Leu Leu Val Thr Val
              -15              -10              -5
Ser Asp Cys Ala Val Ile Thr Gly Ala Cys Glu Arg Asp Val Gln Cys
              1              5              10
Gly Ala Gly Thr Cys Cys Ala Ile Ser Leu Trp Leu Arg Gly Leu Arg
      15              20              25
Met Cys Thr Pro Leu Gly Arg Glu Gly Glu Glu Cys His Pro Gly Ser
30              35              40              45
His Lys Ile Pro Phe Phe Arg Lys Arg Lys His His Thr Cys Pro Cys
              50              55              60
Leu Pro Asn Leu Leu Cys Ser Arg Phe Pro Asp Gly Arg Tyr Arg Cys
              65              70              75
Ser Met Asp Leu Lys Asn Ile Asn Phe
      80              85

```

<210> 195

<211> 102

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -36...-1

<220>

<221> UNSURE

<222> 21

<223> Xaa = Ile, Met

<220>

<221> UNSURE

<222> 52

<223> Xaa = Ile, Ser

<400> 195

```

Met Ser Asn Thr Thr Val Pro Asn Ala Pro Gln Ala Asn Ser Asp Ser
  -35          -30          -25
Met Val Gly Tyr Val Leu Gly Pro Phe Phe Leu Ile Thr Leu Val Gly
-20          -15          -10          -5
Val Val Val Ala Val Val Met Tyr Val Gln Lys Lys Lys Arg Val Asp
              1              5              10
Arg Leu Arg His His Leu Leu Pro Xaa Tyr Ser Tyr Asp Pro Ala Glu
      15              20              25
Glu Leu His Glu Ala Glu Gln Gln Leu Leu Ser Asp Met Gly Asp Pro
      30              35              40
Lys Val Val His Gly Trp Gln Xaa Gly Tyr Gln His Lys Arg Met Pro
45          50          55          60
Leu Leu Asp Val Lys Thr
              65

```

<210> 196

<211> 137

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -18..-1

<400> 196

```

Met Lys Ala Leu Ile Val Leu Gly Leu Val Leu Leu Ser Val Thr Val
          -15          -10          -5
Gln Gly Lys Val Phe Glu Arg Cys Glu Leu Ala Arg Thr Leu Lys Arg
      1              5              10
Leu Gly Met Asp Gly Tyr Arg Gly Ile Ser Leu Ala Asn Trp Met Cys
15          20          25          30
Leu Ala Lys Trp Glu Ser Gly Tyr Asn Thr Arg Ala Thr Asn Tyr Asn
          35          40          45
Ala Gly Asp Arg Ser Thr Asp Tyr Gly Ile Phe Gln Ile Asn Ser Arg
          50          55          60
Tyr Trp Cys Asn Asp Gly Lys Thr Pro Gly Ala Val Asn Ala Cys His
      65          70          75
Leu Ser Cys Ser Gly Trp His Gly Glu Ile Val Val Lys Thr Glu Met
      80          85          90
Ser Val Ser Met Phe Lys Val Val Glu Cys Asn Ser Arg Ile Phe Leu
95          100          105          110
Leu Gln Leu Ile Leu Ser Leu Ser His
              115

```

<210> 197

<211> 101

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -29..-1

<400> 197

```

Met Arg Glu Glu Lys Lys Pro Phe Glu Arg Glu Arg Glu Ser Val Cys
          -25          -20          -15
Val Cys Met Cys Val Phe Ser Thr Gln Gly Ala Leu Gly Glu Met Ala
          -10          -5              1
Ala His Phe Ile Asp Glu Lys Leu Arg Pro Ser Glu Gly Asn Gly His

```

5	10	15														
Arg	Gly	Thr	Leu	Asp	Ser	Leu	Ser	Ser	Asp	Gln	Glu	Ser	Tyr	Ile	Pro	
20					25					30					35	
Ser	Thr	Ala	Asp	Pro	Thr	Gln	Ala	Gly	Pro	Glu	Leu	Leu	His	Lys	Asn	
				40					45					50		
Leu	Pro	Val	Thr	Ser	Arg	Ser	Gln	Pro	Leu	Pro	Ser	Asp	Leu	Ala	Ile	
			55					60					65			
Pro	Ala	Ala	Ala	Leu												
			70													

<210> 198
 <211> 128
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -26..-1

<220>
 <221> UNSURE
 <222> 49
 <223> Xaa = Glu,Lys

<400> 198																
Met	Cys	Trp	Leu	Arg	Ala	Trp	Gly	Gln	Ile	Leu	Leu	Pro	Val	Phe	Leu	
-25						-20					-15					
Ser	Leu	Phe	Leu	Ile	Gln	Leu	Leu	Ile	Ser	Phe	Ser	Glu	Asn	Gly	Phe	
-10					-5				1				5			
Ile	His	Ser	Pro	Arg	Asn	Asn	Gln	Lys	Pro	Arg	Asp	Gly	Asn	Glu	Glu	
			10				15					20				
Glu	Cys	Ala	Val	Lys	Lys	Ser	Cys	Gln	Leu	Cys	Thr	Glu	Asp	Lys	Lys	
		25				30					35					
Cys	Val	Trp	Cys	Ser	Glu	Glu	Lys	Ala	Cys	Xaa	Lys	Tyr	Cys	Phe	Pro	
	40					45				50						
Tyr	Phe	Gly	Cys	Arg	Phe	Ser	Ser	Ile	Tyr	Trp	Leu	Asn	Cys	Lys	Val	
55					60				65					70		
Asp	Met	Phe	Gly	Ile	Met	Met	Leu	Leu	Leu	Ile	Ala	Val	Leu	Ile	Thr	
				75				80						85		
Gly	Phe	Val	Trp	Tyr	Cys	Cys	Ala	Tyr	His	Phe	Tyr	Leu	Gln	Asp	Ile	
			90				95						100			

<210> 199
 <211> 142
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -24..-1

<220>
 <221> UNSURE
 <222> 90
 <223> Xaa = Asp,Glu

<220>
 <221> UNSURE
 <222> 80
 <223> Xaa = Leu,Arg,Trp

<220>
 <221> UNSURE

<222> 79

<223> Xaa = Lys,Asn

<400> 199

```

Met Arg Leu Ser Trp Phe Arg Val Leu Thr Val Leu Ser Ile Cys Leu
              -20              -15              -10
Ser Ala Val Ala Thr Ala Thr Gly Ala Glu Gly Lys Arg Lys Leu Gln
              -5              1              5
Ile Gly Val Lys Lys Arg Val Asp His Cys Pro Ile Lys Ser Arg Lys
    10              15              20
Gly Asp Val Leu His Met His Tyr Thr Gly Lys Leu Glu Asp Gly Thr
25              30              35              40
Glu Phe Asp Ser Ser Leu Pro Gln Asn Gln Pro Phe Val Phe Ser Leu
              45              50              55
Gly Thr Gly Gln Val Ile Lys Gly Trp Asp Gln Gly Leu Leu Gly Met
              60              65              70
Cys Glu Gly Glu Lys Arg Xaa Xaa Val Ile Pro Ser Glu Leu Gly Tyr
              75              80              85
Gly Xaa Arg Gly Ala Pro Pro Lys Ile Pro Gly Gly Ala Thr Leu Val
    90              95              100
Phe Glu Val Glu Leu Leu Lys Ile Glu Arg Arg Thr Glu Leu
105              110              115

```

<210> 200

<211> 70

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -24..-1

```

<223> Von Heijne matrix
      score 6.54728084640774
      seq ILLTRPACLGSWA/EI

```

<400> 200

```

Met Ala Ser Ser Thr Ser Val Ser Thr Gly Gln Ile Leu Leu Thr Arg
              -20              -15              -10
Pro Ala Cys Leu Gly Ser Trp Ala Glu Ile Arg Ser Pro Val Arg Thr
              -5              1              5
Ile Ser Ile Ala Ser Asp Phe Pro Thr Ala Arg Val Ser Leu Trp Val
    10              15              20
Pro Pro Ala Pro Gly Met Val Pro Ile Lys Ile Ser Gly Cys Ala Asn
25              30              35              40
Trp Ala Phe Ser Pro Ala
              45

```

<210> 201

<211> 113

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -23..-1

<220>

<221> UNSURE

<222> 8

<223> Xaa = His,Lys,Asn,Gln

<400> 201

```

Met Ala Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly Ala Leu Ser Gly

```

140

```

      -20      -15      -10
Ala Ala Ala Leu Gly Phe Ala Ser Tyr Gly Ala His Gly Ala Xaa Phe
      -5      1      5
Pro Asp Ala Tyr Gly Lys Glu Leu Phe Asp Lys Ala Asn Lys His His
10      15      20      25
Phe Leu His Ser Leu Ala Leu Leu Gly Val Pro His Cys Arg Lys Pro
      30      35      40
Leu Trp Ala Gly Leu Leu Leu Ala Ser Gly Thr Thr Leu Phe Cys Thr
      45      50      55
Ser Phe Tyr Tyr Gln Ala Leu Ser Gly Asp Pro Ser Ile Gln Thr Leu
      60      65      70
Ala Pro Ala Gly Gly Thr Leu Leu Leu Leu Gly Trp Leu Ala Leu Ala
      75      80      85
Leu
90

```

```

<210> 202
<211> 113
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> SIGNAL
<222> -23..-1

```

```

<220>
<221> UNSURE
<222> 8
<223> Xaa = Lys,Gln

```

```

<400> 202
Met Ala Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly Ala Leu Ser Gly
      -20      -15      -10
Ala Ala Ala Leu Gly Phe Ala Ser Tyr Gly Ala His Gly Ala Xaa Phe
      -5      1      5
Pro Asp Ala Tyr Gly Lys Glu Leu Phe Asp Lys Ala Asn Lys His His
10      15      20      25
Phe Leu His Ser Leu Ala Leu Leu Gly Val Pro His Cys Arg Lys Pro
      30      35      40
Leu Trp Ala Gly Leu Leu Leu Ala Ser Gly Thr Thr Leu Phe Cys Thr
      45      50      55
Ser Phe Tyr Tyr Gln Ala Leu Ser Gly Asp Pro Ser Ile Gln Thr Leu
      60      65      70
Ala Pro Ala Gly Gly Thr Leu Leu Leu Leu Gly Trp Leu Ala Leu Ala
      75      80      85
Leu
90

```

```

<210> 203
<211> 54
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> SIGNAL
<222> -17..-1
<223> Von Heijne matrix
      score 6.48133858229514
      seq LTFAQLLFATVLG/IA

```

```

<400> 203
Met Phe Arg Arg Leu Thr Phe Ala Gln Leu Leu Phe Ala Thr Val Leu
      -15      -10      -5

```

141

Gly Ile Ala Gly Gly Val Tyr Ile Phe Gln Pro Val Phe Glu Gln Tyr
 1 5 10 15
 Ala Lys Asp Gln Lys Glu Leu Lys Glu Lys Met Gln Leu Val Gln Glu
 20 25 30
 Ser Glu Glu Lys Lys Ser
 35

<210> 204
 <211> 83
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1

<400> 204
 Met Arg Leu Phe Leu Ser Leu Pro Val Leu Val Val Val Leu Ser Ile
 -15 -10 -5 1
 Val Leu Glu Gly Pro Ala Pro Ala Gln Gly Thr Pro Asp Val Ser Ser
 5 10 15
 Ala Leu Asp Lys Leu Lys Glu Phe Gly Asn Thr Leu Glu Asp Lys Ala
 20 25 30
 Arg Glu Leu Ile Ser Arg Ile Lys Gln Ser Glu Leu Ser Ala Lys Met
 35 40 45
 Arg Glu Trp Phe Ser Glu Thr Phe Gln Lys Val Lys Asp Lys Leu Lys
 50 55 60 65
 Ile Asp Ser

<210> 205
 <211> 98
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -43..-1

<400> 205
 Met Phe Arg Ile Glu Gly Leu Ala Pro Lys Leu Asp Pro Glu Glu Met
 -40 -35 -30
 Lys Arg Lys Met Arg Glu Asp Val Ile Ser Ser Ile Arg Asn Phe Leu
 -25 -20 -15
 Ile Tyr Val Ala Leu Leu Arg Val Ser Glu Cys Leu Pro Gly Cys Asp
 -10 -5 1 5
 Cys Asp Thr Ser Gly Glu Leu Thr Asp Gly His Pro Leu Thr Leu Arg
 10 15 20
 Gly His Arg Gly Leu Arg Thr Glu Leu Asn Gly Ser Gly Glu Gln Gly
 25 30 35
 Gly Ser Ile Tyr Leu Lys Glu Ile Gly Gln His Met Lys Thr Gly His
 40 45 50
 His Ile
 55

<210> 206
 <211> 82
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -43..-1
 <223> Von Heijne matrix

score 5.65876793443964
seq VVSFALIATLVYA/LF

<400> 206

```
Met Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp
      -40                      -35          -30
Ser Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val
      -25                      -20          -15
Ser Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val
      -10                      -5              1              5
His Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg
              10              15              20
Glu Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val
              25              30              35
Leu His
```

<210> 207

<211> 82

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -43..-1

<400> 207

```
Met Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp
      -40                      -35          -30
Ser Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val
      -25                      -20          -15
Ser Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val
      -10                      -5              1              5
His Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg
              10              15              20
Glu Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val
              25              30              35
Leu His
```

<210> 208

<211> 82

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -43..-1

<220>

<221> UNSURE

<222> 32

<223> Xaa = Ala,Gly

<400> 208

```
Met Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp
      -40                      -35          -30
Ser Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val
      -25                      -20          -15
Ser Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val
      -10                      -5              1              5
His Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg
              10              15              20
Glu Gln Leu Gln Thr Glu Gln Asp Ala Pro Xaa Asp Ser Thr Ala Val
              25              30              35
```

Leu His

<210> 209
 <211> 82
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -43..-1

<400> 209
 Met Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp
 -40 -35 -30
 Ser Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val
 -25 -20 -15
 Ser Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val
 -10 -5 1 5
 His Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg
 10 15 20
 Glu Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val
 25 30 35

Leu His

<210> 210
 <211> 99
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -28..-1

<220>
 <221> UNSURE
 <222> 2,16
 <223> Xaa = Phe,Ser

<400> 210
 Met Thr Ser Ala Leu Thr Gln Gly Leu Glu Arg Ile Pro Asp Gln Leu
 -25 -20 -15
 Gly Tyr Leu Val Leu Ser Glu Gly Ala Val Leu Ala Ser Xaa Gly Asp
 -10 -5 1
 Leu Glu Asn Asp Glu Gln Ala Ala Ser Ala Ile Xaa Glu Leu Val Ser
 5 10 15 20
 Thr Ala Cys Gly Phe Arg Leu His Arg Gly Met Asn Val Pro Phe Lys
 25 30 35
 Arg Leu Ser Val Val Phe Gly Glu His Thr Leu Leu Val Thr Val Ser
 40 45 50
 Gly Gln Arg Val Phe Val Val Lys Arg Gln Asn Arg Gly Arg Glu Pro
 55 60 65
 Ile Asp Val
 70

<210> 211
 <211> 99
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -28..-1

<220>

<221> UNSURE

<222> 16

<223> Xaa = Phe,Ser

<400> 211

```

Met Thr Ser Ala Leu Thr Gln Gly Leu Glu Arg Ile Pro Asp Gln Leu
      -25                -20                -15
Gly Tyr Leu Val Leu Ser Glu Gly Ala Val Leu Ala Ser Ser Gly Asp
      -10                -5                1
Leu Glu Asn Asp Glu Gln Ala Ala Ser Ala Ile Xaa Glu Leu Val Ser
5      10      15      20
Thr Ala Cys Gly Phe Arg Leu His Arg Gly Met Asn Val Pro Phe Lys
      25      30      35
Arg Leu Ser Val Val Phe Gly Glu His Thr Leu Leu Val Thr Val Ser
      40      45      50
Gly Gln Arg Val Phe Val Val Lys Arg Gln Asn Arg Gly Arg Glu Pro
      55      60      65
Ile Asp Val
      70

```

<210> 212

<211> 186

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -15..-1

<220>

<221> UNSURE

<222> 37

<223> Xaa = Cys,Gly

<220>

<221> UNSURE

<222> 11

<223> Xaa = Cys,Trp

<220>

<221> UNSURE

<222> 149

<223> Xaa = Gly,Arg

<400> 212

```

Met His Val Met Ala Ala Ser Met Ala Arg Gly Gly Val Ser Ala Arg
-15      -10      -5      1
Val Leu Leu Gln Ala Ala Arg Gly Thr Xaa Trp Asn Arg Pro Gly Gly
      5      10      15
Thr Ser Gly Ser Gly Glu Gly Val Ala Leu Gly Thr Thr Arg Lys Phe
      20      25      30
Gln Ala Thr Xaa Ser Arg Pro Ala Gly Glu Glu Asp Ala Gly Gly Pro
      35      40      45
Glu Arg Pro Gly Asp Val Val Asn Val Val Phe Val Asp Arg Ser Gly
50      55      60      65
Gln Arg Ile Pro Val Ser Gly Arg Val Gly Asp Asn Val Leu His Leu
      70      75      80
Ala Gln Arg His Gly Val Asp Leu Glu Gly Ala Cys Glu Ala Ser Leu
      85      90      95
Ala Cys Ser Thr Cys His Val Tyr Val Ser Glu Asp His Leu Asp Leu
      100      105      110
Leu Pro Pro Pro Glu Glu Arg Glu Asp Asp Met Leu Asp Met Ala Pro

```

145

115		120		125													
Leu	Leu	Gln	Glu	Asn	Ser	Arg	Leu	Gly	Cys	Gln	Ile	Val	Leu	Thr	Pro		
130					135					140					145		
Glu	Leu	Glu	Xaa	Ala	Glu	Phe	Thr	Leu	Pro	Lys	Ile	Thr	Arg	Asn	Phe		
				150					155					160			
Tyr	Val	Asp	Gly	His	Val	Pro	Lys	Pro	His								
			165					170									

<210> 213
 <211> 123
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -48..-1

<220>
 <221> UNSURE
 <222> 75
 <223> Xaa = Ala,Asp

<400> 213																	
Met	Glu	Ala	Leu	Gly	Lys	Leu	Lys	Gln	Phe	Asp	Ala	Tyr	Pro	Lys	Thr		
			-45					-40					-35				
Leu	Glu	Asp	Phe	Arg	Val	Lys	Thr	Cys	Gly	Gly	Ala	Thr	Val	Thr	Ile		
		-30					-25					-20					
Val	Ser	Gly	Leu	Leu	Met	Leu	Leu	Leu	Phe	Leu	Ser	Glu	Leu	Gln	Tyr		
	-15					-10					-5						
Tyr	Leu	Thr	Thr	Glu	Val	His	Pro	Glu	Leu	Tyr	Val	Asp	Lys	Ser	Arg		
1			5					10					15				
Gly	Asp	Lys	Leu	Lys	Ile	Asn	Ile	Asp	Val	Leu	Phe	Pro	His	Met	Pro		
		20						25					30				
Cys	Ala	Tyr	Leu	Ser	Ile	Asp	Ala	Met	Asp	Val	Ala	Gly	Glu	Gln	Gln		
	35						40					45					
Leu	Asp	Val	Glu	His	Asn	Leu	Phe	Lys	Gln	Arg	Leu	Asp	Lys	Asp	Gly		
	50					55					60						
Ile	Pro	Val	Ser	Ser	Glu	Ala	Glu	Arg	His	Xaa							
65					70					75							

<210> 214
 <211> 77
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -29..-1

<400> 214																	
Met	Ala	Ala	Thr	Asp	Phe	Val	Gln	Glu	Met	Arg	Ala	Val	Gly	Glu	Arg		
			-25						-20					-15			
Leu	Leu	Leu	Lys	Leu	Gln	Arg	Leu	Pro	Gln	Ala	Glu	Pro	Val	Glu	Ile		
		-10						-5					1				
Val	Ala	Phe	Ser	Val	Ile	Ile	Leu	Phe	Thr	Ala	Thr	Val	Leu	Leu	Leu		
	5					10					15						
Leu	Leu	Ile	Ala	Cys	Ser	Cys	Cys	Cys	Thr	His	Cys	Cys	Cys	Pro	Glu		
20					25					30					35		
Arg	Arg	Gly	Arg	Lys	Val	Gln	Val	Gln	Pro	Thr	Pro	Pro					
				40					45								

<210> 215
 <211> 107

```
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -27..-1
```

<400> 215															
Met	Leu	Phe	Glu	Glu	Ala	Leu	Pro	Leu	Ser	Cys	Ser	Asp	Pro	Val	Leu
		-25					-20					-15			
Ser	Thr	Leu	Ser	Leu	Val	Gln	Phe	Ser	Pro	Ser	Gly	Arg	Thr	Gln	Asp
	-10					-5					1				5
Leu	Leu	Ser	Pro	Gly	Val	Glu	Asn	Leu	Ser	Val	Leu	Asp	Val	Ser	Pro
				10					15					20	
Leu	Gly	Leu	Ala	Cys	Cys	Leu	Leu	Thr	Leu	Thr	Met	Ser	Cys	Pro	Gly
			25					30					35		
Pro	Asp	Pro	Pro	Glu	Gly	Pro	Gly	Thr	Gln	Arg	Val	Trp	Gln	Gly	Ala
		40					45					50			
Leu	Arg	Ile	Leu	Gln	Leu	Pro	Gly	Ala	Pro	Asp	Gly	Val	Ser	Pro	Tyr
	55					60					65				
Gln	Pro	Val	Trp	Ser	Arg	Thr	Pro	Asp	Leu	Lys					
70					75					80					

```
<210> 216
<211> 112
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> SIGNAL  
<222> -39..-1
```

<400> 216																
Met	Leu	Glu	His	Leu	Ser	Ser	Leu	Pro	Thr	Gln	Met	Asp	Tyr	Lys	Gly	
				-35					-30					-25		
Gln	Lys	Leu	Ala	Glu	Gln	Met	Phe	Gln	Gly	Ile	Ile	Leu	Phe	Ser	Ala	
				-20					-15					-10		
Ile	Val	Gly	Phe	Ile	Tyr	Gly	Tyr	Val	Ala	Glu	Gln	Phe	Gly	Trp	Thr	
		-5					1					5				
Val	Tyr	Ile	Val	Met	Ala	Gly	Phe	Ala	Phe	Ser	Cys	Leu	Leu	Thr	Leu	
10						15					20					
Pro	Pro	Trp	Pro	Ile	Tyr	Arg	Arg	His	Pro	Leu	Lys	Trp	Leu	Pro	Val	
				30					35					40		
Gln	Ala	Gln	Thr	Thr	Arg	Asn	Gln	Gly	Lys	Glu	Lys	Leu	Arg	Gly	Met	
			45						50				55			
Leu	Lys	Ile	Ile	Glu	Val	Phe	Met	Ile	Gln	His	Leu	Leu	Leu	Phe	Leu	
		60					65					70				

```
<210> 217
<211> 71
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -23..-1
<223> Von Heijne matrix
      score 5.918755252201
      seq SALMLLP+CRPVL+T/SV
```

<400> 217
Met Met Asn Asn Gly Leu Leu Gln Gln Pro Ser Ala Leu Met Leu Leu
-20 -15 -10

147

```

Pro Cys Arg Pro Val Leu Thr Ser Val Ala Leu Asn Ala Asn Phe Val
    -5              1              5
Ser Trp Lys Ser Arg Thr Lys Tyr Thr Ile Thr Pro Val Lys Met Arg
10              15              20              25
Lys Ser Gly Gly Arg Asp His Thr Gly Ala Gly Asn Val Arg Ser Asn
    30              35              40
Ser Arg Pro Ser Ile Gln Arg
    45

```

<210> 218
 <211> 71
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -23..-1
 <223> Von Heijne matrix
 score 5.918755252201
 seq SALMLLPCRPLVT/SV

```

<400> 218
Met Met Asn Asn Gly Leu Leu Gln Gln Pro Ser Ala Leu Met Leu Leu
    -20              -15              -10
Pro Cys Arg Pro Val Leu Thr Ser Val Ala Leu Asn Ala Asn Phe Val
    -5              1              5
Ser Trp Lys Ser Arg Thr Lys Tyr Thr Ile Thr Pro Val Lys Met Arg
10              15              20              25
Lys Ser Gly Gly Arg Asp His Thr Gly Ala Gly Asn Val Arg Ser Asn
    30              35              40
Ser Arg Pro Ser Ile Gln Arg
    45

```

<210> 219
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1
 <223> Von Heijne matrix
 score 4.68058603039206
 seq VLAGSLLGPTSR/AA

```

<400> 219
Met Ala Val Leu Ala Gly Ser Leu Leu Gly Pro Thr Ser Arg Ser Ala
-15              -10              -5              1
Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro Arg Ala Trp Leu Gly Phe
    5              10              15
Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln Ala Arg Gly Lys Ala
    20              25              30
Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys Arg Lys Asn Lys His
    35              40              45
Gly Trp Val Arg Arg Leu Ser Thr Pro Ala Gly Xaa Gln Val Ile Leu
50              55              60              65
Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser His
    70              75

```

<210> 220
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1

<220>
 <221> UNSURE
 <222> 56
 <223> Xaa = Asn, Ser

<400> 220
 Met Ala Val Leu Ala Gly Ser Leu Leu Gly Pro Thr Ser Arg Ser Ala
 -15 -10 -5 1
 Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro Arg Ala Trp Leu Gly Phe
 5 10 15
 Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln Ala Arg Gly Lys Ala
 20 25 30
 Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys Arg Lys Asn Lys His
 35 40 45
 Gly Trp Val Arg Arg Leu Xaa Thr Pro Ala Gly Val Gln Val Ile Leu
 50 55 60 65
 Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser His
 70 75

<210> 221
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1

<220>
 <221> UNSURE
 <222> 56
 <223> Xaa = Asn, Ser

<220>
 <221> UNSURE
 <222> 62
 <223> Xaa = Lys, Gln, Arg

<400> 221
 Met Ala Val Leu Ala Gly Ser Leu Leu Gly Pro Thr Ser Arg Ser Ala
 -15 -10 -5 1
 Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro Arg Ala Trp Leu Gly Phe
 5 10 15
 Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln Ala Arg Gly Lys Ala
 20 25 30
 Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys Arg Lys Asn Lys His
 35 40 45
 Gly Trp Val Arg Arg Leu Xaa Thr Pro Ala Gly Val Xaa Val Ile Leu
 50 55 60 65
 Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser His
 70 75

<210> 222
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>

<221> SIGNAL
<222> -15..-1

<220>
<221> UNSURE
<222> 61
<223> Xaa = * ,Glu,Leu,Val

<220>
<221> UNSURE
<222> 62
<223> Xaa = Ala,Glu,Pro,Gln

<220>
<221> UNSURE
<222> 56
<223> Xaa = Asn,Ser

<220>
<221> UNSURE
<222> -6
<223> Xaa = Asp,Gly

<220>
<221> UNSURE
<222> 6
<223> Xaa = Gly,Ser

<220>
<221> UNSURE
<222> 63
<223> Xaa = Phe,Val

<400> 222
Met Ala Val Leu Ala Gly Ser Leu Leu Xaa Pro Thr Ser Arg Ser Ala
-15 -10 -5 1
Ala Leu Leu Gly Xaa Arg Trp Leu Gln Pro Arg Ala Trp Leu Gly Phe
5 10 15
Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln Ala Arg Gly Lys Ala
20 25 30
Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys Arg Lys Asn Lys His
35 40 45
Gly Trp Val Arg Arg Leu Xaa Thr Pro Ala Gly Xaa Xaa Xaa Ile Leu
50 55 60 65
Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser His
70 75

<210> 223
<211> 87
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -39..-1

<400> 223
Met Lys Val Leu Leu Leu Thr Gly Leu Gly Ala Leu Phe Phe Ala Tyr
-35 -30 -25
Tyr Trp Asp Asp Asn Phe Asp Pro Ala Ser Leu Gln Gly Ala Arg Val
-20 -15 -10
Leu Leu Thr Gly Ala Asn Ala Gly Val Gly Glu Glu Leu Ala Tyr His
-5 1 5

150

Tyr Ala Arg Leu Gly Ser His Leu Val Leu Thr Ala His Thr Glu Ala
 10 15 20 25
 Leu Leu Gln Lys Ala Arg Trp Leu Thr Leu Val Val Ser Thr Leu Gly
 30 35 40
 Gly Arg Gly Lys Trp Ile Thr
 45

<210> 224
 <211> 125
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1

<220>
 <221> UNSURE
 <222> 100
 <223> Xaa = Asn,Thr

<220>
 <221> UNSURE
 <222> 30
 <223> Xaa = Ile,Met

<400> 224
 Met Lys Leu Leu Thr His Asn Leu Leu Ser Ser His Val Arg Gly Val
 -15 -10 -5 1
 Gly Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala Thr Glu Val Arg Ile
 5 10 15
 Cys Pro Val Glu Phe Asn Pro Asn Phe Val Ala Arg Xaa Ile Pro Lys
 20 25 30
 Val Glu Trp Ser Ala Phe Leu Glu Ala Ala Asp Asn Leu Arg Leu Ile
 35 40 45
 Gln Val Pro Lys Gly Pro Val Glu Gly Tyr Glu Glu Asn Glu Glu Phe
 50 55 60 65
 Leu Arg Thr Met His His Leu Leu Leu Glu Val Glu Val Ile Glu Gly
 70 75 80
 Thr Leu Gln Cys Pro Glu Ser Gly Arg Met Phe Pro Ile Ser Arg Gly
 85 90 95
 Ile Pro Xaa Met Leu Leu Ser Glu Glu Glu Thr Glu Ser
 100 105 110

<210> 225
 <211> 93
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -22..-1

<400> 225
 Met Lys Ile Ser Val Ala Ala Ile Pro Phe Phe Leu Leu Ile Thr Ile
 -20 -15 -10
 Ala Leu Gly Thr Lys Thr Glu Ser Ser Ser Arg Gly Pro Tyr His Pro
 -5 1 5 10
 Ser Glu Cys Cys Phe Thr Tyr Thr Thr Tyr Lys Ile Pro Arg Gln Arg
 15 20 25
 Ile Met Asp Tyr Tyr Glu Thr Asn Ser Gln Cys Ser Lys Pro Gly Ile
 30 35 40
 Val Phe Ile Thr Lys Arg Gly His Ser Val Cys Thr Asn Pro Ser Asp

151

Lys Trp Val Gln Asp Tyr Ile Lys Asp Met Lys Glu Asn
45 50 55
60 65 70

```
<210> 226
<211> 72
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -42..-1
```

<400> 226																
Met	Asn	Arg	Leu	Ala	Gly	Val	Gly	Trp	Arg	Val	Asp	Tyr	Thr	Leu	Ser	
		-40					-35					-30				
Ser	Ser	Leu	Leu	Gln	Ser	Val	Glu	Glu	Pro	Met	Val	His	Leu	Arg	Leu	
	-25					-20					-15					
Glu	Val	Ala	Ala	Ala	Pro	Gly	Thr	Pro	Ala	Gln	Pro	Val	Ala	Met	Ser	
-10					-5					1				5		
Leu	Ser	Ala	Asp	Lys	Phe	Gln	Val	Leu	Leu	Ala	Glu	Leu	Lys	Gln	Ala	
		10						15					20			
Gln	Thr	Leu	Met	Ser	Ser	Leu	Gly									
		25					30									

```
<210> 227
<211> 84
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -20..-1
<223> Von Heijne matrix
      score 4.37180298395146
      seq SLLLSLPPHQGLT/FS
```

[illegible]

```
<210> 228
<211> 54
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -19..-1
```

<400> 228
Met Lys Phe Asn Phe Val Leu Leu Ile Leu Ser Lys Val Leu Asp Asp
 -15 -10 -5
Thr Phe Gln Ser Val Lys Lys Trp Leu Asn Tyr Phe Gln Phe Thr Leu

152

```

      1           5           10
Arg Asn Gly Leu Met Trp Pro Gly Ala Val Ala His Ala Cys Asn Pro
      15           20           25
Ser Thr Gly Ser Arg Leu
      30           35

```

<210> 229
 <211> 54
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19..-1
 <223> Von Heijne matrix
 score 4.24406450728052
 seq LLILSKVLDDTFQ/SV

```

<400> 229
Met Lys Phe Asn Phe Val Leu Leu Ile Leu Ser Lys Val Leu Asp Asp
      -15           -10           -5
Thr Phe Gln Ser Val Lys Lys Trp Leu Asn Tyr Phe Gln Phe Thr Leu
      1           5           10
Arg Asn Gly Leu Met Trp Pro Gly Ala Val Ala His Ala Cys Asn Pro
      15           20           25
Ser Thr Gly Ser Arg Leu
      30           35

```

<210> 230
 <211> 54
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19..-1
 <223> Von Heijne matrix
 score 4.24406450728052
 seq LLILSKVLDDTFQ/SV

```

<400> 230
Met Lys Phe Asn Phe Val Leu Leu Ile Leu Ser Lys Val Leu Asp Asp
      -15           -10           -5
Thr Phe Gln Ser Val Lys Lys Trp Leu Asn Tyr Phe Gln Phe Thr Leu
      1           5           10
Arg Asn Gly Leu Met Trp Pro Gly Ala Val Ala His Ala Cys Asn Pro
      15           20           25
Ser Thr Gly Ser Arg Leu
      30           35

```

<210> 231
 <211> 91
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19..-1

```

<400> 231
Met Tyr Gly Lys Ile Ile Phe Val Leu Leu Leu Ser Glu Ile Val Ser
      -15           -10           -5
Ile Ser Ala Leu Ser Thr Thr Glu Val Ala Met His Thr Ser Thr Ser

```

153

			1				5				10				
Ser	Ser	Val	Thr	Lys	Ser	Tyr	Ile	Ser	Ser	Gln	Thr	Asn	Gly	Glu	Met
	15					20					25				
Gly	Gln	Leu	Val	His	Arg	Phe	Thr	Val	Pro	Ala	Pro	Val	Val	Ile	Ile
30					35					40					45
Leu	Ile	Ile	Leu	Cys	Val	Met	Ala	Gly	Ile	Ile	Gly	Thr	Ile	Leu	Leu
				50					55					60	
Ile	Ser	Tyr	Ser	Ile	Arg	Arg	Leu	Ile	Lys	Ala					
			65				70								

```
<210> 232
<211> 114
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -32..-1
```

[illegible]

```
<210> 233
<211> 75
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -35..-1
```

<400> 233															
Met	Ala	Pro	Glu	Val	Leu	Pro	Lys	Pro	Arg	Met	Arg	Gly	Leu	Leu	Ala
-35					-30					-25					-20
Arg	Arg	Leu	Arg	Asn	His	Met	Ala	Val	Ala	Phe	Val	Leu	Ser	Leu	Gly
				-15					-10					-5	
Val	Ala	Ala	Leu	Tyr	Lys	Phe	Arg	Val	Ala	Asp	Gln	Arg	Lys	Asn	Ala
			1				5					10			
Tyr	Ala	Asp	Phe	Tyr	Arg	Asn	Tyr	Asp	Val	Met	Lys	Asp	Phe	Glu	Glu
	15					20					25				
Met	Arg	Lys	Ala	Gly	Ile	Phe	Gln	Ser	Val	Lys					
30					35					40					

```
<210> 234
<211> 108
<212> PRT
<213> Homo sapiens
```

<220>

<221> SIGNAL
<222> -40..-1

<400> 234

Met	Gln	Lys	Phe	Arg	Lys	Met	Ser	Glu	Thr	His	His	Ser	Val	Ile	Ser
-40					-35					-30					-25
Val	Lys	Ala	Ser	Ser	Pro	Trp	Leu	Ser	Ser	Ser	Val	Thr	Ala	Pro	Ser
			-20					-15						-10	
Met	Val	Ala	Pro	Val	Thr	Phe	Ala	Ser	Ile	Val	Glu	Glu	Glu	Leu	Gln
		-5						1			5				
Gln	Glu	Ala	Ala	Leu	Ile	Arg	Ser	Arg	Glu	Lys	Pro	Leu	Ala	Leu	Ile
10						15				20					
Gln	Ile	Glu	Glu	His	Ala	Ile	Gln	Asp	Leu	Leu	Val	Phe	Tyr	Glu	Ala
25				30					35					40	
Phe	Gly	Asn	Pro	Glu	Glu	Phe	Val	Ile	Val	Glu	Arg	Thr	Pro	Gln	Gly
			45						50					55	
Pro	Leu	Ala	Val	Pro	Met	Trp	Asn	Lys	His	Gly	Cys				
		60						65							

<210> 235
<211> 127
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -47..-1

<400> 235

Met	Phe	Thr	Pro	Leu	Thr	Val	Lys	Tyr	Ala	Tyr	Tyr	Asp	Thr	Glu	Arg
		-45					-40					-35			
Ile	Gly	Val	Asp	Leu	Ile	Met	Lys	Thr	Cys	Phe	Ser	Pro	Asn	Arg	Val
-30					-25						-20				
Ile	Gly	Leu	Ser	Ser	Asp	Leu	Gln	Gln	Val	Gly	Gly	Ala	Ser	Ala	Arg
-15					-10					-5				1	
Ile	Gln	Asp	Ala	Leu	Ser	Thr	Val	Leu	Gln	Tyr	Ala	Glu	Asp	Val	Leu
		5					10					15			
Ser	Gly	Lys	Val	Ser	Ala	Asp	Asn	Thr	Val	Gly	Arg	Phe	Leu	Met	Ser
	20					25					30				
Leu	Val	Asn	Gln	Val	Pro	Lys	Ile	Val	Pro	Asp	Asp	Phe	Glu	Thr	Met
35					40					45					
Leu	Asn	Ser	Asn	Ile	Asn	Asp	Leu	Leu	Met	Val	Thr	Tyr	Leu	Ala	Asn
50				55					60					65	
Leu	Thr	Gln	Ser	Gln	Ile	Ala	Leu	Asn	Glu	Lys	Leu	Val	Asn	Leu	
			70						75					80	

<210> 236
<211> 85
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -32..-1

<400> 236

Met	Lys	Ala	Ala	Leu	Glu	Asp	Thr	Leu	Ala	Glu	Thr	Glu	Ala	Arg	Phe
		-30					-25					-20			
Gly	Ala	Gln	Leu	Ala	His	Ile	Gln	Ala	Leu	Ile	Ser	Gly	Ile	Glu	Ala
-15					-10						-5				
Gln	Leu	Gly	Asp	Val	Arg	Ala	Asp	Ser	Glu	Arg	Gln	Asn	Gln	Glu	Tyr
1			5					10					15		
Gln	Arg	Leu	Met	Asp	Ile	Lys	Ser	Arg	Leu	Glu	Gln	Glu	Ile	Ala	Thr

155

Tyr Arg Ser Leu Leu Glu Gly Gln Glu Asp His Tyr Asn Asn Leu Ser
 20 25 30
 35 40 45
 Ala Ser Lys Val Leu
 50

<210> 237
 <211> 101
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -25..-1

<400> 237
 Met Val Asp Arg Glu Leu Ala Asp Ile His Glu Asp Ala Lys Thr Cys
 -25 -20 -15 -10
 Leu Val Leu Cys Ser Arg Val Leu Ser Val Ile Ser Val Lys Glu Ile
 -5 1 5
 Lys Thr Gln Leu Ser Leu Gly Arg His Pro Ile Ile Ser Asn Trp Phe
 10 15 20
 Asp Tyr Ile Pro Ser Thr Arg Tyr Lys Asp Pro Cys Glu Leu Leu His
 25 30 35
 Leu Cys Arg Leu Thr Ile Arg Asn Gln Leu Leu Thr Asn Asn Met Leu
 40 45 50 55
 Pro Asp Gly Ile Phe Ser Leu Leu Ile Pro Ala Arg Leu Gln Asn Tyr
 60 65 70
 Leu Asn Leu Glu Ile
 75

<210> 238
 <211> 68
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

<400> 238
 Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn Ser His Cys Phe Ala
 -45 -40 -35
 Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr Ile His Ala Ser Arg
 -30 -25 -20
 Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly Gly Ser Ser Ser Glu
 -15 -10 -5 1
 Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn Pro Gly Pro Gly Pro
 5 10 15
 Arg Trp Cys Ser
 20

<210> 239
 <211> 68
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

<400> 239
 Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn Ser His Cys Phe Ala

156

```

      -45          -40          -35
Arg  Glu  Lys  Glu  Gly  Asp  Lys  Pro  Ser  Ala  Thr  Ile  His  Ala  Ser  Arg
    -30          -25          -20
Thr  Met  Pro  Leu  Leu  Ser  Leu  His  Ser  Arg  Gly  Gly  Ser  Ser  Ser  Glu
-15          -10          -5          1
Ser  Ser  Arg  Val  Ser  Leu  His  Cys  Cys  Ser  Asn  Pro  Gly  Pro  Gly  Pro
          5          10          15
Arg  Trp  Cys  Ser
          20

```

<210> 240
 <211> 68
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1
 <223> Von Heijne matrix
 score 3.72491712201476
 seq PLLSLHSRGGSSS/ES

```

<400> 240
Met  Val  Lys  Ser  Ser  Leu  Gln  Arg  Ile  Leu  Asn  Ser  His  Cys  Phe  Ala
      -45          -40          -35
Arg  Glu  Lys  Glu  Gly  Asp  Lys  Pro  Ser  Ala  Thr  Ile  His  Ala  Ser  Arg
    -30          -25          -20
Thr  Met  Pro  Leu  Leu  Ser  Leu  His  Ser  Arg  Gly  Gly  Ser  Ser  Ser  Glu
-15          -10          -5          1
Ser  Ser  Arg  Val  Ser  Leu  His  Cys  Cys  Ser  Asn  Pro  Gly  Pro  Gly  Pro
          5          10          15
Arg  Trp  Cys  Ser
          20

```

<210> 241
 <211> 68
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1
 <223> Von Heijne matrix
 score 3.72491712201476
 seq PLLSLHSRGGSSS/ES

<220>
 <221> unsure
 <222> 12
 <223> Xaa = Lys,Asn

```

<400> 241
Met  Val  Lys  Ser  Ser  Leu  Gln  Arg  Ile  Leu  Asn  Ser  His  Cys  Phe  Ala
      -45          -40          -35
Arg  Glu  Lys  Glu  Gly  Asp  Lys  Pro  Ser  Ala  Thr  Ile  His  Ala  Ser  Arg
    -30          -25          -20
Thr  Met  Pro  Leu  Leu  Ser  Leu  His  Ser  Arg  Gly  Gly  Ser  Ser  Ser  Glu
-15          -10          -5          1
Ser  Ser  Arg  Val  Ser  Leu  Xaa  Xaa  Cys  Ser  Xaa  Pro  Gly  Pro  Gly  Pro
          5          10          15
Arg  Trp  Cys  Ser
          20

```

<210> 242
 <211> 129
 <212> PRT
 <213> Homo sapiens

 <220>
 <221> SIGNAL
 <222> -16..-1
 <223> Von Heijne matrix
 score 3.61791418904325
 seq LLLLLLRALRRGPG/PG

<220>
 <221> unsure
 <222> 92
 <223> Xaa = His,Gln

<220>
 <221> unsure
 <222> 84
 <223> Xaa = Phe,Tyr

<400> 242
 Met Ala Val Leu Leu Leu Leu Leu Arg Ala Leu Arg Arg Gly Pro Gly
 -15 -10 -5
 Pro Gly Pro Arg Pro Leu Trp Gly Pro Gly Pro Ala Trp Ser Pro Gly
 1 5 10 15
 Phe Pro Ala Arg Pro Gly Arg Gly Arg Pro Tyr Met Ala Ser Arg Pro
 20 25 30
 Pro Gly Asp Leu Ala Glu Ala Gly Gly Arg Ala Leu Gln Ser Leu Gln
 35 40 45
 Leu Arg Leu Leu Thr Pro Thr Phe Glu Gly Ile Asn Gly Leu Leu Leu
 50 55 60
 Lys Gln His Leu Val Gln Asn Pro Val Arg Leu Trp Gln Leu Leu Gly
 65 70 75 80
 Gly Thr Phe Xaa Phe Asn Thr Ser Arg Leu Lys Xaa Lys Asn Lys Glu
 85 90 95
 Lys Asp Lys Ser Lys Gly Lys Ala Pro Glu Glu Asp Glu Gly Ile Phe
 100 105 110
 Ile

<210> 243
 <211> 129
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -16..-1
 <223> Von Heijne matrix
 score 3.61791418904325
 seq LLLLLLRALRRGPG/PG

<400> 243
 Met Ala Val Leu Leu Leu Leu Leu Arg Ala Leu Arg Arg Gly Pro Gly
 -15 -10 -5
 Pro Gly Pro Arg Pro Leu Trp Gly Pro Gly Pro Ala Trp Ser Pro Gly
 1 5 10 15
 Phe Pro Ala Arg Pro Gly Arg Gly Arg Pro Tyr Met Ala Ser Arg Pro
 20 25 30
 Pro Gly Asp Leu Ala Glu Ala Gly Gly Arg Ala Leu Gln Ser Leu Gln
 35 40 45
 Leu Arg Leu Leu Thr Pro Thr Phe Glu Gly Ile Asn Gly Leu Leu Leu

158

50 55 60
 Lys Gln His Leu Val Gln Asn Pro Val Arg Leu Trp Gln Leu Leu Gly
 65 70 75 80
 Gly Thr Phe Tyr Phe Asn Thr Ser Arg Leu Lys Gln Lys Asn Lys Glu
 85 90 95
 Lys Asp Lys Ser Lys Gly Lys Ala Pro Glu Glu Asp Glu Gly Ile Phe
 100 105 110
 Ile

<210> 244
 <211> 109
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -33..-1

<220>
 <221> UNSURE
 <222> 43
 <223> Xaa = Asp,Gly,Asn,Ser

<400> 244
 Met Ala Ala Ser Gly Ala Pro Arg Ile Leu Val Asp Leu Leu Lys Leu
 -30 -25 -20
 Asn Val Ala Pro Leu Ala Val Phe Gln Met Leu Lys Ser Met Cys Ala
 -15 -10 -5
 Gly Gln Arg Leu Ala Ser Glu Pro Gln Asp Pro Ala Ala Val Ser Leu
 1 5 10 15
 Pro Thr Ser Ser Val Pro Glu Thr Arg Gly Arg Asn Lys Gly Ser Ala
 20 25 30
 Ala Leu Gly Gly Ala Leu Ala Leu Ala Glu Arg Xaa Ser Arg Glu Gly
 35 40 45
 Ser Ser Gln Arg Met Pro Arg Gln Pro Ser Ala Thr Arg Leu Pro Lys
 50 55 60
 Gly Gly Gly Pro Gly Lys Ser Pro Thr Arg Gly Ser Thr
 65 70 75

<210> 245
 <211> 103
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -36..-1

<400> 245
 Met Ala Asn Leu Phe Ile Arg Lys Met Val Asn Pro Leu Leu Tyr Leu
 -35 -30 -25
 Arg Arg His Thr Val Lys Pro Arg Ala Leu Ser Thr Phe Leu Phe Gly
 -20 -15 -10 -5
 Ser Ile Arg Gly Ala Pro Val Ala Val Glu Pro Gly Ala Ala Val
 1 5 10
 Arg Ser Leu Leu Ser Pro Gly Leu Leu Pro His Leu Leu Pro Ala Leu
 15 20 25
 Gly Phe Lys Asn Lys Thr Val Leu Lys Lys Arg Cys Lys Asp Cys Tyr
 30 35 40
 Leu Val Lys Arg Arg Gly Arg Trp Tyr Val Tyr Cys Lys Thr His Pro
 45 50 55 60
 Arg His Lys Gln Arg Gln Met
 65

<210> 246
 <211> 86
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -36..-1

<400> 246
 Met Gly Pro Ala Leu Arg Ser Leu Gln Val Lys Lys Gly Thr Glu His
 -35 -30 -25
 Ala Asp Pro Leu Pro Phe Pro Ser Val Ser Leu Ser Gly Phe Thr Val
 -20 -15 -10 -5
 Gly Thr Leu Ser Glu Thr Ser Thr Gly Gly Pro Ala Thr Pro Thr Trp
 1 5 10
 Lys Glu Cys Pro Ile Cys Lys Glu Arg Phe Pro Ala Glu Ser Asp Lys
 15 20 25
 Asp Ala Leu Glu Asp His Met Asp Gly His Phe Phe Phe Ser Thr Gln
 30 35 40
 Asp Pro Phe Thr Phe Glu
 45 50

<210> 247
 <211> 52
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -30..-1

<400> 247
 Met Ser Ser Cys Gly Ile Val Gly Ser Ser Val Ser Phe Gln Leu Asp
 -30 -25 -20 -15
 Ala Val Lys Leu Leu Leu Lys Met Val Ser Ser Ala Thr Thr Glu Arg
 -10 -5 1
 Cys Cys Asn Gly Ser Ala Asn Phe His Lys Asn Leu Cys Ala Thr Gly
 5 10 15
 Ile Lys Asn Phe
 20

<210> 248
 <211> 219
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -40..-1
 <223> Von Heijne matrix
 score 6.4143157541603
 seq ITMSATVVDAVNA/AP

<220>
 <221> unsure
 <222> 13
 <223> Xaa = Glu,Gln

<220>
 <221> unsure
 <222> -39

<223> Xaa = Ile,Asn

<400> 248

```

Pro Xaa Arg Lys Ile Thr Arg Leu Ala Trp Ile Leu Ser Asp Gly Leu
-40                               -35           -30           -25
Gly Ala Ser Cys Gly Gly Ala Gly Gly Gly Gln Ile Thr Met Ser Ala
-20                               -15           -10
Thr Val Val Asp Ala Val Asn Ala Ala Pro Leu Ser Gly Ser Lys Glu
-5                               1             5
Met Ser Leu Glu Xaa Pro Lys Lys Met Thr Arg Glu Asp Trp Arg Lys
10                               15           20
Lys Lys Glu Leu Glu Glu Gln Arg Lys Leu Gly Asn Ala Pro Ala Glu
25                               30           35           40
Val Asp Glu Glu Gly Lys Asp Ile Asn Pro His Ile Pro Gln Tyr Ile
45                               50           55
Ser Ser Val Pro Trp Tyr Ile Asp Pro Ser Lys Arg Pro Thr Leu Lys
60                               65           70
His Gln Arg Pro Gln Pro Glu Lys Gln Lys Gln Phe Ser Ser Ser Gly
75                               80           85
Glu Trp Tyr Lys Arg Gly Val Lys Glu Asn Ser Ile Ile Thr Lys Tyr
90                               95           100
Arg Lys Gly Ala Cys Glu Asn Cys Gly Ala Met Thr His Lys Lys Lys
105                              110          115           120
Asp Cys Phe Glu Arg Pro Arg Arg Val Gly Ala Lys Phe Thr Gly Thr
125                              130          135
Asn Ile Ala Pro Asp Glu His Val Gln Pro Gln Leu Met Phe Asp Tyr
140                              145          150
Asp Gly Lys Arg Asp Arg Trp Asn Gly Tyr Asn Pro Glu Glu His Met
155                              160          165
Lys Ile Val Glu Glu Tyr Ala Lys Val Asp Leu
170                              175

```

<210> 249

<211> 215

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -36..-1

```

<223> Von Heijne matrix
      score 5.52542399647426
      seq ITMSATVXDAVNA/AP

```

<400> 249

```

Ile Thr Arg Leu Ala Trp Ile Leu Xaa Asp Gly Leu Gly Ala Ser Cys
-35                               -30           -25
Gly Gly Ala Gly Gly Gly Gln Ile Thr Met Ser Ala Thr Val Xaa Asp
-20                               -15           -10           -5
Ala Val Asn Ala Ala Pro Leu Ser Gly Ser Lys Glu Met Ser Leu Glu
1                               5             10
Glu Pro Lys Lys Met Thr Arg Glu Asp Trp Arg Lys Lys Lys Glu Leu
15                               20           25
Glu Glu Gln Arg Lys Leu Gly Asn Ala Pro Ala Glu Val Asp Glu Glu
30                               35           40
Gly Lys Asp Ile Asn Pro His Ile Pro Gln Tyr Ile Ser Ser Val Pro
45                               50           55           60
Trp Tyr Ile Asp Pro Ser Lys Arg Pro Thr Leu Lys His Gln Arg Pro
65                               70           75
Gln Pro Glu Lys Gln Lys Gln Phe Ser Ser Ser Gly Glu Trp Tyr Lys
80                               85           90
Arg Gly Val Lys Glu Asn Ser Ile Ile Thr Lys Tyr Arg Lys Gly Ala
95                               100          105

```

161

Cys Glu Asn Cys Gly Ala Met Thr His Lys Lys Lys Asp Cys Phe Glu
 110 115 120
 Arg Pro Arg Arg Val Gly Ala Lys Phe Thr Gly Thr Asn Ile Ala Pro
 125 130 135 140
 Asp Glu His Val Gln Pro Gln Leu Met Phe Asp Tyr Asp Gly Lys Arg
 145 150 155
 Asp Arg Trp Asn Gly Tyr Asn Pro Glu Glu His Met Lys Ile Val Glu
 160 165 170
 Glu Tyr Ala Lys Val Asp Leu
 175

<210> 250
 <211> 214
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -36..-1

<220>
 <221> UNSURE
 <222> 120
 <223> Xaa = Glu,Lys

<400> 250
 Ile Thr Arg Leu Ala Trp Ile Leu Ser Asp Gly Leu Gly Ala Ser Cys
 -35 -30 -25
 Gly Gly Ala Gly Gly Gly Gln Ile Thr Met Ser Ala Thr Val Val Asp
 -20 -15 -10 -5
 Ala Val Asn Ala Ala Pro Leu Ser Gly Ser Lys Glu Met Ser Leu Glu
 1 5 10
 Glu Pro Lys Lys Met Thr Arg Glu Asp Trp Arg Lys Lys Lys Glu Leu
 15 20 25
 Glu Glu Gln Arg Lys Leu Gly Asn Ala Pro Ala Glu Val Asp Glu Glu
 30 35 40
 Gly Lys Asp Ile Asn Pro His Ile Pro Gln Tyr Ile Ser Ser Val Pro
 45 50 55 60
 Trp Tyr Ile Asp Pro Ser Lys Arg Pro Thr Leu Lys His Gln Arg Pro
 65 70 75
 Gln Pro Glu Lys Gln Lys Gln Phe Ser Ser Ser Gly Glu Trp Tyr Lys
 80 85 90
 Arg Gly Val Lys Glu Asn Ser Ile Ile Thr Lys Tyr Arg Lys Gly Ala
 95 100 105
 Cys Glu Asn Cys Gly Ala Met Thr His Lys Lys Xaa Asp Cys Phe Glu
 110 115 120
 Arg Pro Arg Arg Val Gly Ala Lys Phe Thr Gly Thr Asn Ile Ala Pro
 125 130 135 140
 Asp Glu His Val Gln Pro Gln Leu Met Phe Asp Tyr Asp Gly Lys Arg
 145 150 155
 Asp Arg Trp Asn Gly Tyr Asn Pro Glu Glu His Met Lys Ile Val Glu
 160 165 170
 Glu Tyr Ala Lys Val Asp
 175

<210> 251
 <211> 147
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1

<220>

<221> UNSURE

<222> -9

<223> Xaa = Glu,Lys

<400> 251

```

Met Asp Val Leu Val Ser Xaa Cys Ser Ala Arg Leu Leu Gln Gln Glu
-15          -10          -5          1
Glu Glu Ile Lys Ser Leu Thr Ala Glu Ile Asp Arg Leu Lys Asn Cys
      5          10          15
Gly Cys Leu Gly Ala Ser Pro Asn Leu Glu Gln Leu Gln Glu Glu Asn
      20          25          30
Leu Lys Leu Lys Tyr Arg Leu Asn Ile Leu Arg Lys Ser Leu Gln Ala
      35          40          45
Glu Arg Asn Lys Pro Thr Lys Asn Met Ile Asn Ile Ile Ser Arg Leu
50          55          60          65
Gln Glu Val Phe Gly His Ala Ile Lys Ala Ala Tyr Pro Asp Leu Glu
      70          75          80
Asn Pro Pro Leu Leu Val Thr Pro Ser Gln Gln Ala Lys Phe Gly Asp
      85          90          95
Tyr Gln Cys Asn Ser Ala Met Gly Ile Ser Gln Val Met Tyr Cys His
      100          105          110
Asp Ser Trp Leu Phe Asp Phe Phe Lys Tyr Tyr Tyr His His Cys His
      115          120          125
Leu Gln Lys
130

```

<210> 252

<211> 117

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -43..-1

```

<223> Von Heijne matrix
      score 4.10872570449087
      seq LGQVAAAAAANA/AS

```

<400> 252

```

Leu Ser Pro Ala Gln Ala Ser Gly Ile Trp Cys Arg Asp Ser Ala Phe
      -40          -35          -30
Ser Gly Trp Ser Glu Arg Gly Ser Ser Ser Arg Arg Gly Cys Leu Gly
      -25          -20          -15
Gln Val Ala Ala Ala Ala Ala Ala Asn Ala Ala Ser Pro Pro Pro
      -10          -5          1          5
Ser Ala Asn Ser Leu Gly Ser Gly Gly Arg Cys Lys Leu Arg Ala Pro
      10          15          20
Thr Ser Arg Pro Ser Gln Ser Arg Pro Arg Ser Leu Pro Gln Ala Arg
      25          30          35
Phe Pro Pro Ser Pro Leu Pro Pro Ser Pro Ala Gly Thr Ser Cys Thr
      40          45          50
Cys Tyr Cys Phe Leu Leu Ile Asn Ala Asp Leu Ile Lys Trp Ser Pro
      55          60          65
Lys Lys Lys Lys Lys
70

```

<210> 253

<211> 88

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -36..-1

<400> 253

Met	Ala	Ser	Val	Val	Pro	Val	Lys	Asp	Lys	Lys	Leu	Leu	Glu	Val	Lys
	-35					-30					-25				
Leu	Gly	Glu	Leu	Pro	Ser	Trp	Ile	Leu	Met	Arg	Asp	Phe	Ser	Pro	Ser
-20					-15					-10					-5
Gly	Ile	Phe	Gly	Ala	Phe	Gln	Arg	Gly	Tyr	Tyr	Arg	Tyr	Tyr	Asn	Lys
				1				5					10		
Tyr	Ile	Asn	Val	Lys	Lys	Gly	Ser	Ile	Ser	Gly	Ile	Thr	Met	Val	Leu
	15					20						25			
Ala	Cys	Tyr	Val	Leu	Phe	Ser	Tyr	Ser	Phe	Ser	Tyr	Lys	His	Leu	Lys
	30					35					40				
His	Glu	Arg	Leu	Arg	Lys	Tyr	His								
45					50										

<210> 254

<211> 62

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -42..-1

<220>

<221> UNSURE

<222> -11

<223> Xaa = Glu,Lys

<400> 254

Met	Asp	Pro	Asn	Cys	Ser	Cys	Ala	Ala	Ala	Gly	Val	Ser	Cys	Thr	Cys
		-40					-35					-30			
Ala	Ser	Ser	Cys	Lys	Cys	Lys	Glu	Cys	Lys	Cys	Thr	Ser	Cys	Lys	Xaa
-25						-20					-15				
Ser	Cys	Cys	Ser	Cys	Cys	Pro	Val	Gly	Cys	Ala	Lys	Cys	Ala	Gln	Gly
-10					-5				1					5	
Cys	Ile	Cys	Lys	Gly	Ala	Ser	Glu	Lys	Cys	Ser	Cys	Cys	Ala		
			10				15						20		

<210> 255

<211> 132

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 118

<223> Xaa = Gly,Val

<400> 255

Leu	Phe	Pro	Ala	Pro	Ala	Pro	Pro	Pro	Ala	Pro	Ala	Phe	Ala	Pro	Pro
1				5					10					15	
Pro	Lys	Val	Pro	Ser	Pro	Glu	Arg	Ser	Ala	Pro	Arg	Val	Pro	Leu	Pro
			20					25					30		
Ser	Pro	Gln	Pro	Ser	Tyr	Pro	Phe	Arg	Pro	Ala	Ala	Ser	Gly	Gly	Thr
		35					40					45			
Pro	Pro	Pro	Ala	Cys	Leu	Pro	Pro	Ala	Gln	Pro	Cys	Gln	Gly	Ser	Pro
	50					55					60				
Ala	Met	Asn	Leu	Phe	Arg	Phe	Leu	Gly	Asp	Leu	Ser	His	Leu	Leu	Ala
65					70				75						80

164

Ile Ile Leu Leu Leu Leu Lys Ile Trp Lys Ser Arg Ser Cys Ala Ala
 85 90 95
 His Pro Gln Leu Pro Leu Ser Phe Cys Leu Ser Val Cys Leu Ser Val
 100 105 110
 Ser Leu Ser Leu Ser Xaa Ser Leu Ser Leu Ser Phe Ser Val Ser Lys
 115 120 125
 Lys Lys Lys Lys
 130

<210> 256
 <211> 163
 <212> PRT
 <213> Homo sapiens

<400> 256
 Pro His Arg Leu Thr Leu Pro Thr Leu Gln Ser Val Thr Phe Arg Cys
 1 5 10 15
 Pro Ser Arg Ser Glu Lys Leu Gly Lys Asn Met Val Ser Ser Phe Arg
 20 25 30
 Val Ser Glu Leu Gln Val Leu Leu Gly Phe Ala Gly Arg Asn Lys Ser
 35 40 45
 Gly Arg Lys His Asp Leu Leu Met Arg Ala Leu His Leu Leu Lys Ser
 50 55 60
 Gly Cys Ser Pro Ala Val Gln Ile Lys Ile Arg Glu Leu Tyr Arg Arg
 65 70 75 80
 Arg Tyr Pro Arg Thr Leu Glu Gly Leu Ser Asp Leu Ser Thr Ile Lys
 85 90 95
 Ser Ser Val Phe Ser Leu Asp Gly Gly Ser Ser Pro Val Glu Pro Asp
 100 105 110
 Leu Ala Val Ala Gly Ile His Ser Leu Pro Ser Thr Ser Val Thr Pro
 115 120 125
 His Ser Pro Ser Ser Pro Val Gly Ser Val Leu Leu Gln Asp Thr Lys
 130 135 140
 Pro Thr Phe Glu Met Gln Gln Pro Ser Pro Pro Ile Pro Pro Val His
 145 150 155 160
 Pro Asp Val

<210> 257
 <211> 166
 <212> PRT
 <213> Homo sapiens

<400> 257
 Pro His Arg Leu Thr Leu Pro Thr Leu Gln Ser Val Thr Phe Arg Cys
 1 5 10 15
 Pro Ser Arg Ser Glu Lys Leu Gly Lys Asn Met Val Ser Ser Phe Arg
 20 25 30
 Val Ser Glu Leu Gln Val Leu Leu Gly Phe Ala Gly Arg Asn Lys Ser
 35 40 45
 Gly Arg Lys His Asp Leu Leu Met Arg Ala Leu His Leu Leu Lys Ser
 50 55 60
 Gly Cys Ser Pro Ala Val Gln Ile Lys Ile Arg Glu Leu Tyr Arg Arg
 65 70 75 80
 Arg Tyr Pro Arg Thr Leu Glu Gly Leu Ser Asp Leu Ser Thr Ile Lys
 85 90 95
 Ser Ser Val Phe Ser Leu Asp Gly Gly Ser Ser Pro Val Glu Pro Asp
 100 105 110
 Leu Ala Val Ala Gly Ile His Ser Leu Pro Ser Thr Ser Val Thr Pro
 115 120 125
 His Ser Pro Ser Ser Pro Val Gly Ser Val Leu Leu Gln Asp Thr Lys
 130 135 140
 Pro Thr Phe Glu Met Gln Gln Pro Ser Pro Pro Ile Pro Pro Val His
 145 150 155 160

Pro Asp Val Gln Leu Lys
165

<210> 258
<211> 118
<212> PRT
<213> Homo sapiens

<400> 258
Met Ser Tyr Glu Gln Leu Met Gln Leu Tyr Ser Ala Arg Gln Arg Arg
1 5 10 15
Arg Leu Asn Arg Gly Leu Arg Arg Lys Gln His Ser Leu Leu Lys Arg
20 25 30
Leu Arg Lys Ala Lys Lys Glu Ala Pro Pro Met Glu Lys Pro Glu Val
35 40 45
Val Lys Thr His Leu Arg Asp Met Ile Ile Leu Pro Glu Met Val Gly
50 55 60
Ser Met Val Gly Val Tyr Asn Gly Lys Thr Phe Asn Gln Val Glu Ile
65 70 75 80
Lys Pro Glu Met Ile Gly His Tyr Leu Gly Glu Phe Ser Ile Thr Tyr
85 90 95
Lys Pro Val Lys His Gly Arg Pro Gly Ile Gly Ala Thr His Ser Ser
100 105 110
Arg Phe Ile Pro Leu Lys
115

<210> 259
<211> 101
<212> PRT
<213> Homo sapiens

<220>
<221> unsure
<222> 98
<223> Xaa = Ile,Asn

<220>
<221> unsure
<222> 19
<223> Xaa = Ser,Thr

<400> 259
Arg Thr Ala Gly Arg Gly Arg Pro Ala Val Ala Ser Trp Glu Leu Arg
1 5 10 15
Ala Arg Xaa Cys Ala Glu Asp Pro His Gln Gly Ala Gly Cys Ser Cys
20 25 30
Gly Ser Arg Ala Met Ala Glu Glu Gln Gly Arg Glu Arg Asp Ser Val
35 40 45
Pro Lys Pro Ser Val Leu Phe Leu His Pro Asp Leu Gly Val Gly Gly
50 55 60
Ala Glu Arg Leu Val Leu Leu Pro Phe Pro Thr Glu Arg Arg Leu Leu
65 70 75 80
Phe Cys Thr Ala Pro Arg Ser Leu Asn Ser Leu Pro Leu Lys Lys Lys
85 90 95
Lys Xaa Xaa Lys Ser
100

<210> 260
<211> 100
<212> PRT
<213> Homo sapiens

<400> 260

166

Val Arg Gln Gly Gly Val Gly Pro Gln Trp Pro Ala Gly Ser Tyr Glu
 1 5 10 15
 Arg Gly Ala Cys Ala Glu Asp Pro His Gln Gly Ala Gly Cys Ser Cys
 20 25 30
 Gly Ser Arg Ala Met Ala Glu Glu Gln Gly Arg Glu Arg Asp Ser Val
 35 40 45
 Pro Lys Pro Ser Val Leu Phe Leu His Pro Asp Leu Gly Val Gly Gly
 50 55 60
 Ala Glu Arg Leu Val Leu Leu Pro Phe Pro Thr Glu Arg Arg Leu Leu
 65 70 75 80
 Phe Cys Thr Ala Pro Arg Ser Leu Asn Ser Leu Pro Leu Lys Lys Lys
 85 90 95
 Lys Asn Ala Lys
 100

<210> 261

<211> 147

<212> PRT

<213> Homo sapiens

<220>

<221> unsure

<222> 90

<223> Xaa = Ala,Gly

<400> 261

Met Glu Tyr Val Leu Glu Val Lys Asn Ser Pro Arg His Leu Leu Lys
 1 5 10 15
 Gln Phe Thr Val Cys Asp Val Pro Leu Tyr Asp Ile Cys Asp Tyr Asn
 20 25 30
 Val Ser Arg Asp Arg Cys Gln Glu Leu Gly Cys Cys Xaa Tyr Glu Gly
 35 40 45
 Xaa Cys Tyr Lys Lys Ala Val Pro Ile Tyr Ile His Val Phe Ser Ala
 50 55 60
 Leu Ile Val Ile Lys Ala Gly Ala Phe Val Ile Thr Ile Ile Tyr Arg
 65 70 75 80
 Val Ile Gln Glu Ser Arg Lys Glu Lys Xaa Ile Pro Val Tyr Val Ala
 85 90 95
 Leu Pro Gln Lys Ser Ser Glu Lys Ala Glu Leu Ala Ser Ser Ser Ser
 100 105 110
 Lys Leu Gly Leu Lys Pro Ala Ser Pro Gly Pro Pro Ser Ala Gly Pro
 115 120 125
 Ser Met Lys Ser Asp Glu Asp Lys Asp Asp Val Thr Gly Thr Ile Thr
 130 135 140
 Glu Ala Glu
 145

<210> 262

<211> 147

<212> PRT

<213> Homo sapiens

<400> 262

Met Glu Tyr Val Leu Glu Val Lys Asn Ser Pro Arg His Leu Leu Lys
 1 5 10 15
 Gln Phe Thr Val Cys Asp Val Pro Leu Tyr Asp Ile Cys Asp Tyr Asn
 20 25 30
 Val Ser Arg Asp Arg Cys Gln Glu Leu Gly Cys Cys Phe Tyr Glu Gly
 35 40 45
 Val Cys Tyr Lys Lys Ala Val Pro Ile Tyr Ile His Val Phe Ser Ala
 50 55 60
 Leu Ile Val Ile Ile Ala Gly Ala Phe Val Ile Thr Ile Ile Tyr Arg
 65 70 75 80

167

Val Ile Gln Glu Ser Arg Lys Glu Lys Ala Ile Pro Val Tyr Val Ala
 85 90 95
 Leu Pro Gln Lys Ser Ser Glu Lys Ala Glu Leu Ala Ser Ser Ser Ser
 100 105 110
 Lys Leu Gly Leu Lys Pro Ala Ser Pro Gly Pro Pro Ser Ala Gly Pro
 115 120 125
 Ser Met Lys Ser Asp Glu Asp Lys Asp Asp Val Thr Gly Thr Ile Thr
 130 135 140
 Glu Ala Glu
 145

<210> 263
 <211> 149
 <212> PRT
 <213> Homo sapiens

<400> 263
 Met Arg Tyr Asn Glu Lys Glu Leu Gln Ala Leu Ser Arg Gln Pro Ala
 1 5 10 15
 Glu Met Ala Ala Glu Leu Gly Met Arg Gly Pro Lys Lys Gly Ser Val
 20 25 30
 Leu Lys Arg Arg Leu Val Lys Leu Val Val Asn Phe Leu Phe Tyr Phe
 35 40 45
 Arg Thr Asp Glu Ala Glu Pro Val Gly Ala Leu Leu Leu Glu Arg Cys
 50 55 60
 Arg Val Val Arg Glu Glu Pro Gly Thr Phe Ser Ile Ser Phe Ile Glu
 65 70 75 80
 Asp Pro Glu Arg Lys Tyr His Phe Glu Cys Ser Ser Glu Glu Gln Cys
 85 90 95
 Gln Glu Trp Met Glu Ala Leu Arg Arg Ala Ser Tyr Glu Phe Met Arg
 100 105 110
 Arg Ser Leu Ile Phe Tyr Arg Asn Glu Ile Arg Lys Val Thr Gly Lys
 115 120 125
 Asp Pro Leu Glu Gln Phe Gly Ile Ser Glu Glu Ala Arg Phe Gln Leu
 130 135 140
 Ser Gly Leu Gln Ala
 145

<210> 264
 <211> 103
 <212> PRT
 <213> Homo sapiens

<400> 264
 Met Pro Val Val Pro Ala Leu Gly Arg Pro Arg Trp Ala Asp His Leu
 1 5 10 15
 Arg Ser Gly Val Arg Asp Gln Pro Gly Gln Pro Gly Glu Ala Pro Pro
 20 25 30
 Ser Leu Leu Lys Ile Gln Lys Leu Ala Gly Tyr Gly Gly Gly Cys Leu
 35 40 45
 Trp Ser Gln Leu Leu Gly Arg Leu Arg Arg Glu Asn His Leu Ser Pro
 50 55 60
 Gly Gly Gly Gly Cys Ser Glu Pro Arg Leu Cys His Cys Thr Pro Ala
 65 70 75 80
 Trp Val Thr Glu Gln Asp Ser Ile Ser Lys Ile Glu Lys Ile Ser Gln
 85 90 95
 Leu Pro Lys Lys Lys Lys Lys
 100

<210> 265
 <211> 78
 <212> PRT
 <213> Homo sapiens

<400> 265

```

Met Asp Pro Asn Cys Ser Cys Ala Ala Gly Val Ser Cys Thr Cys Ala
1          5          10          15
Ser Ser Cys Lys Cys Lys Glu Cys Lys Cys Thr Ser Cys Lys Lys Ser
          20          25          30
Cys Cys Ser Trp Cys Thr Gly Pro Thr Leu Val Asp Thr Cys Pro Ala
          35          40          45
Leu Pro Pro Val Cys Leu Ser Gln Arg Ser Ser Gly Met Asn Leu Arg
          50          55          60
Thr His Val Gln Trp Glu Val Arg Pro Pro Leu Asn Ile Gln
65          70          75

```

<210> 266

<211> 92

<212> PRT

<213> Homo sapiens

<400> 266

```

Met Ser Pro Gln Thr His Ser Gln Thr Cys Ile Arg Asn Leu Val Thr
1          5          10          15
Cys Ile Asn Tyr Pro Arg Thr Ser Thr Gly Cys Lys Gly Thr Thr Thr
          20          25          30
Gln Arg Ile Met Glu Pro Val Glu Leu Glu Val Glu Gly Thr Glu Gln
          35          40          45
Asp Asn Ala Lys Thr Cys Gly Ser Leu Gly Arg Gly Asn Glu Asn Thr
          50          55          60
Met Leu Arg Gly Gly Phe Ser Met Asn Thr Thr Val Gly Gln Gly Ile
65          70          75          80
Ser Lys Gln Thr His His Thr Ser Thr Thr Ser Ser
          85          90

```

<210> 267

<211> 51

<212> PRT

<213> Homo sapiens

<400> 267

```

Met Leu Ala Lys Thr Gly Val His His Tyr Ser Gly Asn Asn Ile Glu
1          5          10          15
Leu Gly Thr Ala Cys Gly Lys Tyr Tyr Arg Val Cys Thr Leu Ala Ile
          20          25          30
Ile Asp Pro Gly Asp Ser Asp Ile Ile Arg Ser Met Pro Glu Gln Thr
          35          40          45
Gly Glu Lys
          50

```

<210> 268

<211> 99

<212> PRT

<213> Homo sapiens

<400> 268

```

Met Lys Lys Lys Glu Glu Thr Thr Leu Ser Glu Met Glu Pro Val Glu
1          5          10          15
Pro Gln Tyr Gln Leu Val Asn Ala Glu Ser Thr Ser Pro Phe Leu His
          20          25          30
Cys Leu Arg Glu Val Ile Gly Glu Tyr Ser Val His Glu Phe Ser Leu
          35          40          45
Leu Gly Lys Thr Glu Ser Gln Gly Ile Gly Leu Trp Ile Ala Leu Val
          50          55          60
Val Phe Leu Ser Phe Leu Ile Phe Ser Thr Ser Phe Tyr Ile Ser Asn
65          70          75          80

```

169

Ala Glu Gln Pro Phe Phe Lys Glu Pro Pro Thr Glu Ala Ala Lys Glu
 85 90 95
 Leu Ser Leu

<210> 269
 <211> 128
 <212> PRT
 <213> Homo sapiens

<400> 269
 Met Ala Gly Asn Leu Leu Ser Gly Ala Gly Arg Arg Leu Trp Asp Trp
 1 5 10 15
 Val Pro Leu Ala Cys Arg Ser Phe Ser Leu Gly Val Pro Arg Leu Ile
 20 25 30
 Gly Ile Arg Leu Thr Leu Pro Pro Pro Lys Val Val Asp Arg Trp Asn
 35 40 45
 Glu Lys Arg Ala Met Phe Gly Val Tyr Asp Asn Ile Gly Ile Leu Gly
 50 55 60
 Asn Phe Glu Lys His Pro Lys Glu Leu Ile Arg Gly Pro Ile Trp Leu
 65 70 75 80
 Arg Gly Trp Lys Gly Asn Glu Leu Gln Arg Cys Ile Arg Lys Arg Lys
 85 90 95
 Met Val Gly Ser Arg Met Phe Ala Asp Asp Leu His Asn Leu Asn Lys
 100 105 110
 Arg Ile Arg Tyr Leu Tyr Lys His Phe Asn Arg His Gly Lys Phe Arg
 115 120 125

<210> 270
 <211> 52
 <212> PRT
 <213> Homo sapiens

<400> 270
 Met Phe Leu Thr Leu Ala Asp Thr Cys Lys Leu Arg Gly Met Ser Phe
 1 5 10 15
 Leu Leu Asn Val Tyr Glu Gly Glu Ala Thr Val Ser Ser Val Leu Glu
 20 25 30
 Leu Leu Glu Ser Trp Ile Ile Val Gly Asn Glu Arg Tyr Phe Asp Gly
 35 40 45
 Ile Ser Ser His
 50

<210> 271
 <211> 84
 <212> PRT
 <213> Homo sapiens

<400> 271
 Met Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp
 1 5 10 15
 Val Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly
 20 25 30
 Cys Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp
 35 40 45
 Gly Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu
 50 55 60
 His Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp
 65 70 75 80
 Lys Phe Lys Glu

<210> 272
 <211> 84
 <212> PRT

<213> Homo sapiens

<400> 272

```

Met Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp
1      5      10      15
Val Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly
      20      25      30
Cys Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp
      35      40      45
Gly Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu
      50      55      60
His Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp
65      70      75      80
Lys Phe Lys Glu

```

<210> 273

<211> 84

<212> PRT

<213> Homo sapiens

<400> 273

```

Met Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp
1      5      10      15
Val Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly
      20      25      30
Cys Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp
      35      40      45
Gly Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu
      50      55      60
His Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp
65      70      75      80
Lys Phe Lys Glu

```

<210> 274

<211> 84

<212> PRT

<213> Homo sapiens

<400> 274

```

Met Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp
1      5      10      15
Val Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly
      20      25      30
Cys Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp
      35      40      45
Gly Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu
      50      55      60
His Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp
65      70      75      80
Lys Phe Lys Glu

```

<210> 275

<211> 84

<212> PRT

<213> Homo sapiens

<400> 275

```

Met Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp
1      5      10      15
Val Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly
      20      25      30
Cys Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp
      35      40      45

```

171

Gly Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu
 50 55 60
 His Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp
 65 70 75 80
 Lys Phe Lys Glu

<210> 276
 <211> 80
 <212> PRT
 <213> Homo sapiens

<400> 276
 Met Ala Thr Pro Thr Gln Thr Pro Thr Lys Ala Pro Glu Glu Pro Asp
 1 5 10 15
 Pro Phe Tyr Tyr Asp Tyr Asn Thr Val Gln Thr Val Gly Met Thr Leu
 20 25 30
 Ala Thr Ile Leu Phe Leu Leu Gly Ile Leu Ile Val Ile Ser Lys Lys
 35 40 45
 Val Lys Cys Arg Lys Ala Asp Ser Arg Ser Glu Ser Pro Thr Cys Lys
 50 55 60
 Ser Cys Lys Ser Glu Leu Pro Ser Ser Ala Pro Gly Gly Gly Gly Val
 65 70 75 80

<210> 277
 <211> 127
 <212> PRT
 <213> Homo sapiens

<400> 277
 Met Ser Phe Ser Gly Lys Tyr Gln Leu Gln Ser Gln Glu Asn Phe Glu
 1 5 10 15
 Ala Phe Met Lys Ala Ile Gly Leu Pro Glu Glu Leu Ile Gln Lys Gly
 20 25 30
 Lys Asp Ile Lys Gly Val Ser Glu Ile Val Gln Asn Gly Lys His Phe
 35 40 45
 Lys Phe Thr Ile Thr Ala Gly Ser Lys Val Ile Gln Asn Glu Phe Thr
 50 55 60
 Val Gly Glu Glu Cys Glu Leu Glu Thr Met Thr Gly Glu Lys Val Lys
 65 70 75 80
 Thr Val Val Gln Leu Glu Gly Asp Asn Lys Leu Val Thr Thr Phe Lys
 85 90 95
 Asn Ile Lys Ser Val Thr Glu Leu Asn Gly Asp Ile Ile Thr Asn Thr
 100 105 110
 Met Thr Leu Gly Asp Ile Val Phe Lys Arg Ile Ser Lys Arg Ile
 115 120 125

<210> 278
 <211> 87
 <212> PRT
 <213> Homo sapiens

<400> 278
 Asp Ala Val Ala Val Gly Lys Phe Ser Gln Trp Pro Pro Glu Arg Ser
 1 5 10 15
 Pro Ala Asp Ala Leu Cys Asp Gly Lys Lys Ala Ala Gly Pro Thr Ser
 20 25 30
 Gly Ala Phe Tyr Asn Glu Leu Arg Ala Leu Glu Ser Leu Pro Gly Asn
 35 40 45
 Tyr Ser Ser Ala Ser Thr Val Val Leu Leu Trp Ile Ala Ser Arg Gln
 50 55 60
 Lys Ser Gln Ser Ser Arg Asn Ser Gln Cys Ser Arg Lys Asp Lys Gly
 65 70 75 80
 Gly Lys Gln Glu Glu His Glu

85

<210> 279
 <211> 96
 <212> PRT
 <213> Homo sapiens

<400> 279
 Met Pro Gly Pro Thr Pro Ser Gly Thr Asn Val Gly Ser Ser Gly Arg
 1 5 10 15
 Ser Pro Ser Lys Ala Val Ala Ala Arg Ala Ala Gly Ser Thr Val Arg
 20 25 30
 Gln Arg Lys Asn Ala Ser Cys Gly Thr Arg Ser Ala Gly Arg Thr Thr
 35 40 45
 Ser Ala Gly Thr Gly Gly Met Trp Arg Phe Tyr Thr Glu Asp Ser Pro
 50 55 60
 Gly Leu Lys Val Gly Pro Val Pro Val Leu Val Met Ser Leu Leu Phe
 65 70 75 80
 Ile Ala Ser Val Phe Met Leu His Ile Trp Gly Lys Tyr Thr Arg Ser
 85 90 95

<210> 280
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 280
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45
 Lys Lys Tyr Tyr Glu Lys Met Pro
 50 55

<210> 281
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 281
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45
 Lys Lys Tyr Tyr Glu Lys Met Pro
 50 55

<210> 282
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 282
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45

Lys Lys Tyr Tyr Glu Lys Met Pro
50 55

<210> 283
<211> 56
<212> PRT
<213> Homo sapiens

<400> 283
Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
1 5 10 15
Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
20 25 30
Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
35 40 45
Lys Lys Tyr Tyr Glu Lys Met Pro
50 55

<210> 284
<211> 56
<212> PRT
<213> Homo sapiens

<400> 284
Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
1 5 10 15
Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
20 25 30
Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
35 40 45
Lys Lys Tyr Tyr Glu Lys Met Pro
50 55

<210> 285
<211> 56
<212> PRT
<213> Homo sapiens

<400> 285
Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
1 5 10 15
Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
20 25 30
Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
35 40 45
Lys Lys Tyr Tyr Glu Lys Met Pro
50 55

<210> 286
<211> 56
<212> PRT
<213> Homo sapiens

<400> 286
Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
1 5 10 15
Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
20 25 30
Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
35 40 45
Lys Lys Tyr Tyr Glu Lys Met Pro
50 55

<210> 287
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 287
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45
 Lys Lys Tyr Tyr Glu Lys Met Pro
 50 55

<210> 288
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 288
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45
 Lys Lys Tyr Tyr Glu Lys Met Pro
 50 55

<210> 289
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 289
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45
 Lys Lys Tyr Tyr Glu Lys Met Pro
 50 55

<210> 290
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 290
 Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45
 Lys Lys Tyr Tyr Glu Lys Met Pro
 50 55

<210> 291
 <211> 56
 <212> PRT

<213> Homo sapiens

<400> 291

Met	Ser	Ser	Trp	Ser	Leu	Met	Cys	Met	Gln	Asp	Met	Gly	Asn	Thr	Lys
1				5					10					15	
Ala	Arg	Leu	Leu	Tyr	Glu	Ala	Asn	Leu	Pro	Glu	Asn	Phe	Arg	Arg	Pro
		20					25						30		
Gln	Thr	Asp	Gln	Ala	Val	Glu	Phe	Phe	Ile	Arg	Asp	Lys	Tyr	Glu	Lys
		35					40					45			
Lys	Lys	Tyr	Tyr	Glu	Lys	Met	Pro								
	50					55									

<210> 292

<211> 56

<212> PRT

<213> Homo sapiens

<400> 292

Met	Ser	Ser	Trp	Ser	Leu	Met	Cys	Met	Gln	Asp	Met	Gly	Asn	Thr	Lys
1				5					10					15	
Ala	Arg	Leu	Leu	Tyr	Glu	Ala	Asn	Leu	Pro	Glu	Asn	Phe	Arg	Arg	Pro
		20					25						30		
Gln	Thr	Asp	Gln	Ala	Val	Glu	Phe	Phe	Ile	Arg	Asp	Lys	Tyr	Glu	Lys
		35					40					45			
Lys	Lys	Tyr	Tyr	Glu	Lys	Met	Pro								
	50					55									

<210> 293

<211> 56

<212> PRT

<213> Homo sapiens

<400> 293

Met	Ser	Ser	Trp	Ser	Leu	Met	Cys	Met	Gln	Asp	Met	Gly	Asn	Thr	Lys
1				5					10					15	
Ala	Arg	Leu	Leu	Tyr	Glu	Ala	Asn	Leu	Pro	Glu	Asn	Phe	Arg	Arg	Pro
		20					25						30		
Gln	Thr	Asp	Gln	Ala	Val	Glu	Phe	Phe	Ile	Arg	Asp	Lys	Tyr	Glu	Lys
		35					40					45			
Lys	Lys	Tyr	Tyr	Glu	Lys	Met	Pro								
	50					55									

<210> 294

<211> 56

<212> PRT

<213> Homo sapiens

<400> 294

Met	Ser	Ser	Trp	Ser	Leu	Met	Cys	Met	Gln	Asp	Met	Gly	Asn	Thr	Lys
1				5					10					15	
Ala	Arg	Leu	Leu	Tyr	Glu	Ala	Asn	Leu	Pro	Glu	Asn	Phe	Arg	Arg	Pro
		20					25						30		
Gln	Thr	Asp	Gln	Ala	Val	Glu	Phe	Phe	Ile	Arg	Asp	Lys	Tyr	Glu	Lys
		35					40					45			
Lys	Lys	Tyr	Tyr	Glu	Lys	Met	Pro								
	50					55									

<210> 295

<211> 122

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 113

<223> Xaa = Ala,Asp

<400> 295

```

Met Glu Ala Ser Ala Leu Thr Ser Ser Ala Val Thr Ser Val Ala Lys
1          5          10          15
Val Val Arg Val Ala Ser Gly Ser Ala Val Val Leu Pro Leu Ala Arg
          20          25          30
Ile Ala Thr Val Val Ile Gly Gly Val Val Ala Met Ala Ala Val Pro
          35          40          45
Met Val Leu Ser Ala Met Gly Phe Thr Ala Ala Gly Ile Ala Ser Ser
          50          55          60
Ser Ile Ala Ala Lys Met Met Ser Ala Ala Ala Ile Ala Asn Gly Gly
65          70          75          80
Gly Val Ala Ser Gly Ser Leu Val Ala Thr Leu Gln Ser Leu Gly Ala
          85          90          95
Thr Gly Leu Ser Gly Leu Thr Lys Phe Ile Leu Gly Ser Ile Gly Ser
          100          105          110
Xaa Ile Ala Ala Val Ile Ala Arg Phe Tyr
          115          120

```

<210> 296

<211> 122

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 79

<223> Xaa = Gly,Arg

<400> 296

```

Met Glu Ala Ser Ala Leu Thr Ser Ser Ala Val Thr Ser Val Ala Lys
1          5          10          15
Val Val Arg Val Ala Ser Gly Ser Ala Val Val Leu Pro Leu Ala Arg
          20          25          30
Ile Ala Thr Val Val Ile Gly Gly Val Val Ala Met Ala Ala Val Pro
          35          40          45
Met Val Leu Ser Ala Met Gly Phe Thr Ala Ala Gly Ile Ala Ser Ser
          50          55          60
Ser Ile Ala Ala Lys Met Met Ser Ala Ala Ala Ile Ala Asn Xaa Gly
65          70          75          80
Gly Val Ala Ser Gly Ser Leu Val Ala Thr Leu Gln Ser Leu Gly Ala
          85          90          95
Thr Gly Leu Ser Gly Leu Thr Lys Phe Ile Leu Gly Ser Ile Gly Ser
          100          105          110
Ala Ile Ala Ala Val Ile Ala Arg Phe Tyr
          115          120

```

<210> 297

<211> 74

<212> PRT

<213> Homo sapiens

<400> 297

```

Met Ala Ser Ser Thr Val Pro Val Ser Ala Ala Gly Ser Ala Asn Glu
1          5          10          15
Thr Pro Glu Ile Pro Asp Asn Val Gly Asp Trp Leu Arg Gly Val Tyr
          20          25          30
Arg Phe Ala Thr Asp Arg Asn Asp Phe Arg Arg Asn Leu Ile Leu Asn
          35          40          45
Leu Gly Leu Phe Ala Ala Gly Val Trp Leu Ala Arg Asn Leu Ser Asp

```

50 55 60
 Ile Asp Leu Met Ala Pro Gln Pro Gly Val
 65 70

<210> 298
 <211> 74
 <212> PRT
 <213> Homo sapiens

<400> 298
 Met Ala Ser Ser Thr Val Pro Val Ser Ala Ala Gly Ser Ala Asn Glu
 1 5 10 15
 Thr Pro Glu Ile Pro Asp Asn Val Gly Asp Trp Leu Arg Gly Val Tyr
 20 25 30
 Arg Phe Ala Thr Asp Arg Asn Asp Phe Arg Arg Asn Leu Ile Leu Asn
 35 40 45
 Leu Gly Leu Phe Ala Ala Gly Val Trp Leu Ala Arg Asn Leu Ser Asp
 50 55 60
 Ile Asp Leu Met Ala Pro Gln Pro Gly Val
 65 70

<210> 299
 <211> 147
 <212> PRT
 <213> Homo sapiens

<400> 299
 Met Asp Val Leu Val Ser Glu Cys Ser Ala Arg Leu Leu Gln Gln Glu
 1 5 10 15
 Glu Glu Ile Lys Ser Leu Thr Ala Glu Ile Asp Arg Leu Lys Asn Cys
 20 25 30
 Gly Cys Leu Gly Ala Ser Pro Asn Leu Glu Gln Leu Gln Glu Glu Asn
 35 40 45
 Leu Lys Leu Lys Tyr Arg Leu Asn Ile Leu Arg Lys Ser Leu Gln Ala
 50 55 60
 Glu Arg Asn Lys Pro Thr Lys Asn Met Ile Asn Ile Ile Ser Arg Leu
 65 70 75 80
 Gln Glu Val Phe Gly His Ala Ile Lys Ala Ala Tyr Pro Asp Leu Glu
 85 90 95
 Asn Pro Pro Leu Leu Val Thr Pro Ser Gln Gln Ala Lys Phe Gly Asp
 100 105 110
 Tyr Gln Cys Asn Ser Ala Met Gly Ile Ser Gln Val Met Tyr Cys His
 115 120 125
 Asp Ser Trp Leu Phe Asp Phe Lys Tyr Tyr Tyr His His Cys His
 130 135 140
 Leu Gln Lys
 145

<210> 300
 <211> 97
 <212> PRT
 <213> Homo sapiens

<400> 300
 Met Ser Ser Ala Pro Glu Pro Pro Thr Phe Lys Lys Glu Pro Pro Lys
 1 5 10 15
 Glu Lys Glu Phe Gln Ser Pro Gly Leu Arg Gly Val Arg Thr Thr Thr
 20 25 30
 Leu Phe Arg Ala Val Asn Pro Glu Leu Phe Ile Lys Pro Asn Lys Pro
 35 40 45
 Val Met Ala Phe Gly Leu Val Thr Leu Ser Leu Cys Val Ala Tyr Ile
 50 55 60
 Gly Tyr Leu His Ala Ile Gln Glu Asn Lys Lys Asp Leu Tyr Glu Ala

178

65 70 75 80
 Ile Asp Ser Glu Gly His Ser Tyr Met Arg Arg Lys Thr Ser Lys Trp
 85 90 95

Asp

<210> 301

<211> 98

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 30

<223> Xaa = Lys,Asn

<400> 301

Met Ala Ala Glu Pro Leu Thr Glu Leu Glu Glu Ser Ile Glu Asn Val
 1 5 10 15
 Val Thr Thr Phe Phe Thr Phe Ala Arg Gln Glu Gly Arg Xaa Asp Ser
 20 25 30
 Leu Ser Val Asn Glu Phe Lys Glu Leu Val Thr Gln Gln Leu Pro His
 35 40 45
 Leu Leu Lys Asp Val Gly Ser Leu Asp Glu Lys Met Lys Ser Leu Asp
 50 55 60
 Val Asn Gln Asp Ser Glu Leu Lys Phe Asn Glu Tyr Trp Arg Leu Ile
 65 70 75 80
 Gly Glu Leu Ala Lys Glu Ile Arg Lys Lys Lys Asp Leu Lys Ile Arg
 85 90 95

Lys Lys

<210> 302

<211> 115

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 103

<223> Xaa = Asp,Tyr

<220>

<221> UNSURE

<222> 100

<223> Xaa = Gly,Ser

<400> 302

Met Val Ala Ala Lys Lys Thr Lys Lys Ser Leu Glu Ser Ile Lys Ser
 1 5 10 15
 Arg Leu Gln Leu Val Met Lys Ser Gly Lys Tyr Val Leu Gly Tyr Lys
 20 25 30
 Gln Thr Leu Lys Met Ile Arg Gln Gly Lys Ala Lys Leu Val Ile Leu
 35 40 45
 Ala Asn Asn Cys Pro Ala Leu Arg Lys Ser Glu Ile Glu Tyr Tyr Ala
 50 55 60
 Met Leu Ala Lys Thr Gly Val His His Tyr Ser Gly Asn Asn Ile Glu
 65 70 75 80
 Leu Gly Thr Ala Cys Gly Lys Tyr Tyr Arg Val Cys Thr Leu Ala Ile
 85 90 95
 Ile Asp Pro Xaa Asp Ser Xaa Ile Ile Arg Ser Met Pro Glu Gln Thr
 100 105 110

Gly Glu Lys
 115

<210> 303
 <211> 102
 <212> PRT
 <213> Homo sapiens

<400> 303
 Met Ala Gly Gln Ala Phe Arg Lys Phe Leu Pro Leu Phe Asp Arg Val
 1 5 10 15
 Leu Val Glu Arg Ser Pro Ala Glu Thr Val Thr Lys Gly Gly Ile Met
 20 25 30
 Leu Pro Glu Lys Ser Gln Gly Lys Val Leu Gln Ala Thr Val Val Ala
 35 40 45
 Val Gly Ser Gly Ser Lys Gly Lys Gly Glu Ile Gln Pro Val Ser
 50 55 60
 Val Lys Val Gly Asp Lys Val Leu Leu Pro Glu Tyr Gly Gly Thr Lys
 65 70 75 80
 Val Val Leu Asp Asn Lys Asp Tyr Phe Leu Phe Arg Asp Gly Asp Ile
 85 90 95
 Leu Gly Lys Tyr Val Asp
 100

<210> 304
 <211> 102
 <212> PRT
 <213> Homo sapiens

<400> 304
 Met Ala Gly Gln Ala Phe Arg Lys Phe Leu Pro Leu Phe Asp Arg Val
 1 5 10 15
 Leu Val Glu Arg Ser Ala Ala Glu Thr Val Thr Lys Gly Gly Ile Met
 20 25 30
 Leu Pro Glu Lys Ser Gln Gly Lys Val Leu Gln Ala Thr Val Val Ala
 35 40 45
 Val Gly Ser Gly Ser Lys Gly Lys Gly Glu Ile Gln Pro Val Ser
 50 55 60
 Val Lys Val Gly Asp Lys Val Leu Leu Pro Glu Tyr Gly Gly Thr Lys
 65 70 75 80
 Val Val Leu Asp Asp Lys Asp Tyr Phe Leu Phe Arg Asp Gly Asp Ile
 85 90 95
 Leu Gly Lys Tyr Val Asp
 100

<210> 305
 <211> 116
 <212> PRT
 <213> Homo sapiens

<400> 305
 Met Ser Ala Thr Ala Ala Thr Ala Pro Pro Ala Ala Pro Ala Gly Glu
 1 5 10 15
 Gly Gly Pro Pro Ala Pro Pro Pro Asn Leu Thr Ser Asn Arg Arg Leu
 20 25 30
 Gln Gln Thr Gln Ala Gln Val Asp Glu Val Val Asp Ile Met Arg Val
 35 40 45
 Asn Val Asp Lys Val Leu Glu Arg Asp Gln Lys Leu Ser Glu Leu Asp
 50 55 60
 Asp Arg Ala Asp Ala Leu Gln Ala Gly Pro Ser Gln Phe Glu Thr Ser
 65 70 75 80
 Ala Ala Lys Leu Lys Arg Lys Tyr Trp Trp Lys Asn Leu Lys Met Met
 85 90 95
 Ile Ile Leu Gly Val Ile Cys Ala Ile Ile Leu Ile Ile Ile Ile Val
 100 105 110
 Tyr Phe Ser Thr

115

<210> 306
 <211> 97
 <212> PRT
 <213> Homo sapiens

<400> 306

Met	Ser	Leu	Ser	Leu	Val	Phe	Arg	Ala	Ala	Ser	Tyr	Phe	Lys	Leu	Val
1				5					10					15	
Pro	Phe	His	Ser	Ser	Ser	Ser	Asn	Gln	Phe	Leu	Gln	Pro	Pro	Gly	Trp
			20					25					30		
Val	Val	Leu	Thr	Gln	Thr	Leu	Val	Leu	Leu	His	Phe	Glu	Arg	Phe	Ser
		35					40					45			
Tyr	Gln	Asn	Val	Pro	Lys	Ser	Ala	Gln	Gly	Lys	Gly	Asn	Leu	Gln	Pro
	50					55					60				
Glu	Thr	Asn	Ile	His	Leu	Phe	His	Phe	Leu	Thr	Phe	Pro	Lys	Gln	Ile
65					70					75					80
Ser	Arg	Asn	Leu	Phe	Asn	Ser	Leu	Leu	Cys	Leu	Met	Cys	Leu	Thr	Tyr
				85					90					95	

Phe

<210> 307
 <211> 110
 <212> PRT
 <213> Homo sapiens

<400> 307

Met	Ser	Met	Thr	Asp	Leu	Leu	Asn	Ala	Glu	Asp	Ile	Lys	Lys	Ala	Val
1				5					10					15	
Gly	Ala	Phe	Ser	Ala	Thr	Asp	Ser	Phe	Asp	His	Lys	Lys	Phe	Phe	Gln
			20					25					30		
Met	Val	Gly	Leu	Lys	Lys	Lys	Ser	Ala	Asp	Asp	Val	Lys	Lys	Val	Phe
		35					40					45			
His	Met	Leu	Asp	Lys	Asp	Lys	Ser	Gly	Phe	Ile	Glu	Glu	Asp	Glu	Leu
	50					55					60				
Gly	Phe	Ile	Leu	Lys	Gly	Phe	Ser	Pro	Asp	Ala	Arg	Asp	Leu	Ser	Ala
65					70				75						80
Lys	Glu	Thr	Lys	Met	Leu	Met	Ala	Ala	Gly	Asp	Lys	Asp	Gly	Asp	Gly
				85					90					95	
Lys	Ile	Gly	Val	Asp	Glu	Phe	Ser	Thr	Leu	Val	Ala	Glu	Ser		
			100					105						110	

<210> 308
 <211> 110
 <212> PRT
 <213> Homo sapiens

<400> 308

Met	Ser	Met	Thr	Asp	Leu	Leu	Asn	Ala	Glu	Asp	Ile	Lys	Lys	Ala	Val
1				5					10					15	
Gly	Ala	Phe	Ser	Ala	Thr	Asp	Ser	Phe	Asp	His	Lys	Lys	Phe	Phe	Gln
			20					25					30		
Met	Val	Gly	Leu	Lys	Lys	Lys	Ser	Ala	Asp	Asp	Val	Lys	Lys	Val	Phe
		35					40					45			
His	Met	Leu	Asp	Lys	Asp	Lys	Ser	Gly	Phe	Ile	Glu	Glu	Asp	Glu	Leu
	50					55					60				
Gly	Phe	Ile	Leu	Lys	Gly	Phe	Ser	Pro	Asp	Ala	Arg	Asp	Leu	Ser	Ala
65					70				75						80
Lys	Glu	Thr	Lys	Met	Leu	Met	Ala	Ala	Gly	Asp	Lys	Asp	Gly	Asp	Gly
				85					90					95	
Lys	Ile	Gly	Val	Asp	Glu	Phe	Ser	Thr	Leu	Val	Ala	Glu	Ser		
			100					105						110	

<210> 309
 <211> 110
 <212> PRT
 <213> Homo sapiens

<400> 309
 Met Ser Met Thr Asp Leu Leu Asn Ala Glu Asp Ile Lys Lys Ala Val
 1 5 10 15
 Gly Ala Phe Ser Ala Thr Asp Ser Phe Asp His Lys Lys Phe Phe Gln
 20 25 30
 Met Val Gly Leu Lys Lys Lys Ser Ala Asp Asp Val Lys Lys Val Phe
 35 40 45
 His Met Leu Asp Lys Asp Lys Ser Gly Phe Ile Glu Glu Asp Glu Leu
 50 55 60
 Gly Phe Ile Leu Lys Gly Phe Ser Pro Asp Ala Arg Asp Leu Ser Ala
 65 70 75 80
 Lys Glu Thr Lys Met Leu Met Ala Ala Gly Asp Lys Asp Gly Asp Gly
 85 90 95
 Lys Ile Gly Val Asp Glu Phe Ser Thr Leu Val Ala Glu Ser
 100 105 110

<210> 310
 <211> 112
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 8,33
 <223> Xaa = Asp,Glu

<400> 310
 Met Asp Pro Arg Lys Val Asn Xaa Leu Arg Ala Phe Val Lys Met Cys
 1 5 10 15
 Lys Gln Asp Pro Ser Val Leu His Thr Glu Glu Met Arg Phe Leu Arg
 20 25 30
 Xaa Trp Val Glu Ser Met Gly Gly Lys Val Pro Pro Ala Thr Gln Lys
 35 40 45
 Ala Lys Ser Glu Glu Asn Thr Lys Glu Glu Lys Pro Asp Ser Lys Lys
 50 55 60
 Val Glu Glu Asp Leu Lys Ala Asp Glu Pro Ser Ser Glu Glu Ser Asp
 65 70 75 80
 Leu Glu Ile Asp Lys Glu Gly Val Ile Glu Pro Asp Thr Asp Ala Pro
 85 90 95
 Gln Glu Met Gly Asp Glu Asn Ala Glu Ile Thr Ala Gln His Phe Leu
 100 105 110

<210> 311
 <211> 118
 <212> PRT
 <213> Homo sapiens

<400> 311
 Met Val Leu Leu Glu Ser Glu Gln Phe Leu Thr Glu Leu Thr Arg Leu
 1 5 10 15
 Phe Gln Lys Cys Arg Thr Ser Gly Ser Val Tyr Ile Thr Leu Lys Lys
 20 25 30
 Tyr Asp Gly Arg Thr Lys Pro Ile Pro Lys Lys Gly Thr Val Glu Gly
 35 40 45
 Phe Glu Pro Ala Asp Asn Lys Cys Leu Leu Arg Ala Thr Asp Gly Lys
 50 55 60
 Lys Lys Ile Ser Thr Val Val Ser Ser Lys Glu Val Asn Lys Phe Gln

182															
65					70					75					80
Met	Ala	Tyr	Ser	Asn	Leu	Leu	Arg	Ala	Asn	Met	Asp	Gly	Leu	Lys	Lys
				85					90					95	
Arg	Asp	Lys	Lys	Asn	Glu	Thr	Lys	Lys	Thr	Lys	Ala	Ala	Ala	Ala	Ala
			100					105					110		
Ala	Ala	Thr	Ala	Ala	Gln										
		115													

```
<210> 312
<211> 109
<212> PRT
<213> Homo sapiens
```

<400> 312															
Met	Asp	Ser	Leu	Pro	Ser	Gly	Lys	Ile	His	Arg	Lys	Val	Lys	Ile	Ile
1				5					10					15	
Leu	Gly	Arg	Asn	Arg	Lys	Glu	Asn	Leu	Glu	Pro	Asn	Ala	Glu	Phe	Asp
			20					25					30		
Lys	Arg	Thr	Glu	Phe	Ile	Thr	Gln	Glu	Glu	Asn	Arg	Ile	Cys	Ser	Ser
		35				40						45			
Pro	Val	Gln	Ser	Leu	Leu	Asp	Leu	Phe	Gln	Thr	Ser	Glu	Glu	Lys	Ser
	50					55					60				
Glu	Phe	Leu	Gly	Phe	Thr	Ser	Tyr	Thr	Glu	Lys	Ser	Gly	Ile	Cys	Asn
65				70						75				80	
Val	Leu	Asp	Ile	Trp	Glu	Glu	Glu	Asn	Ser	Asp	Asn	Leu	Leu	Thr	Ala
				85					90					95	
Phe	Phe	Ser	Ser	Pro	Ser	Thr	Ser	Thr	Phe	Thr	Gly	Phe			
			100					105							

```
<210> 313
<211> 103
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> unsure
<222> 99
<223> Xaa = Arg,Thr
```

```
<220>
<221> unsure
<222> 103
<223> Xaa = His,Asn
```

[illegible]

<210>	314
<211>	72
<212>	PRT

<213> Homo sapiens

<400> 314

```

Met Ser Gly Asp Gly Ala Thr Glu Gln Ala Ala Glu Tyr Val Ser Glu
1          5          10          15
Lys Val Lys Lys Ala Glu Lys Lys Leu Glu Glu Asn Pro Tyr Asp Leu
          20          25          30
Asp Ala Trp Ser Ile Leu Ile Arg Glu Ala Gln Asn Gln Pro Ile Asp
          35          40          45
Lys Ala Arg Lys Thr Tyr Glu Arg Leu Val Ala Gln Phe Pro Ser Ser
          50          55          60
Gly Arg Phe Trp Lys Leu Tyr Ile
65          70

```

<210> 315

<211> 106

<212> PRT

<213> Homo sapiens

<400> 315

```

Met Leu Lys Ser Asn Gly Glu Arg Arg Ser Arg Asn Ala Leu Pro Ala
1          5          10          15
Val Tyr Ala Arg Lys Met Ala Ala Ser Gln Gln Gln Ala Ser Ala Ala
          20          25          30
Ser Ser Ala Ala Gly Val Ser Gly Pro Ser Ser Ala Gly Gly Pro Gly
          35          40          45
Pro Gln Gln Gln Pro Gln Pro Pro Ala Gln Leu Val Gly Pro Ala Gln
          50          55          60
Ser Gly Leu Leu Gln Gln Gln Gln Gln Asp Phe Asp Pro Val Gln Arg
65          70          75          80
Tyr Lys Met Leu Ile Pro Gln Leu Lys Glu Ser Leu Gln Val Ile Gly
          85          90          95
Leu Lys Gln Arg Glu Ala Asn Trp Ile Trp
          100          105

```

<210> 316

<211> 86

<212> PRT

<213> Homo sapiens

<400> 316

```

Met Ala Glu Asp Met Glu Thr Lys Ile Lys Asn Tyr Lys Thr Ala Pro
1          5          10          15
Phe Asp Ser Arg Phe Pro Asn Gln Asn Gln Thr Arg Asn Cys Trp Gln
          20          25          30
Asn Tyr Leu Asp Phe His Arg Cys Gln Lys Ala Met Thr Ala Lys Gly
          35          40          45
Gly Asp Ile Ser Val Cys Glu Trp Tyr Gln Arg Val Tyr Gln Ser Leu
          50          55          60
Cys Pro Thr Ser Trp Val Thr Asp Trp Asp Glu Gln Arg Ala Glu Gly
65          70          75          80
Thr Phe Pro Gly Lys Ile
          85

```

<210> 317

<211> 105

<212> PRT

<213> Homo sapiens

<400> 317

```

Met Cys Val Gly Leu Leu Met Val Ser Ile Leu Thr Asp Ile Asn Leu
1          5          10          15
Ser Asn Leu Val Ala Val Gln Tyr Ser Pro Asp Tyr Cys Asn Phe Arg

```


184

20	25	30
Lys Arg Ser Asp	Lys Asn Gln Asp	Ala Ser Thr Phe Cys His Asn Cys
35	40	45
Asn Gln Phe His Leu Cys	Leu Gln Tyr His His Lys Ile Val Leu Pro	
50	55	60
Trp Ser Val Leu Ala	Ile Leu Ser Gln Leu Phe Leu His Ile Thr Ser	
65	70	75
Arg Ile Arg Pro Thr	Ile Tyr Lys Ser Asn Lys Pro Lys Ser Ile Glu	
85	90	95
Ile Leu Ile Gly Phe Val Leu Asn Leu		
100	105	

<210> 318
 <211> 101
 <212> PRT
 <213> Homo sapiens

<400> 318
 Met Glu Arg Pro Asp Lys Ala Ala Leu Asn Ala Leu Gln Pro Pro Glu
 1 5 10 15
 Phe Arg Asn Glu Ser Ser Leu Ala Ser Thr Leu Lys Thr Leu Leu Phe
 20 25 30
 Phe Thr Ala Leu Met Ile Thr Val Pro Ile Gly Leu Tyr Phe Thr Thr
 35 40 45
 Lys Ser Tyr Ile Phe Glu Gly Ala Leu Gly Met Ser Asn Arg Asp Ser
 50 55 60
 Tyr Phe Tyr Ala Ala Ile Val Ala Val Val Ala Val His Val Val Leu
 65 70 75 80
 Ala Leu Phe Val Tyr Val Ala Trp Asn Glu Gly Ser Arg Gln Trp Arg
 85 90 95
 Glu Gly Lys Gln Asp
 100

<210> 319
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 29
 <223> Xaa = Gly,Arg

<220>
 <221> UNSURE
 <222> 28
 <223> Xaa = Lys,Asn,Arg,Ser

<400> 319
 Met Phe Arg Asp Phe Gly Arg Arg Leu Gln Arg Asp Leu Lys Arg Val
 1 5 10 15
 Val Asp Ala Arg Leu Arg Leu Ser Glu Glu Leu Xaa Xaa Gly Arg Ile
 20 25 30
 Lys Pro Lys Pro Val Glu Val Gln Val Val Thr His His Met Gln Arg
 35 40 45
 Tyr Ala Val Trp Phe Gly Gly Ser Met Leu Ala Ser Thr Pro Glu Phe
 50 55 60
 Phe Gln Val Cys His Thr Lys Lys Asp Tyr Glu Glu Tyr Gly Pro Ser
 65 70 75 80
 Ile Cys Arg His Asn Pro Val Phe Gly Val Met Ser
 85 90

<210> 320

<211> 92
 <212> PRT
 <213> Homo sapiens

<400> 320

```

Met Phe Arg Asp Phe Gly Arg Arg Leu Gln Arg Asp Leu Lys Arg Val
1      5      10      15
Val Asp Ala Arg Leu Arg Leu Ser Glu Glu Leu Ser Gly Gly Arg Ile
      20      25      30
Lys Pro Lys Pro Val Glu Val Gln Val Val Thr His His Met Gln Arg
      35      40      45
Tyr Ala Val Trp Phe Gly Gly Ser Met Leu Ala Ser Thr Pro Glu Phe
      50      55      60
Phe Gln Val Cys His Thr Lys Lys Asp Tyr Glu Glu Tyr Gly Pro Ser
65      70      75      80
Ile Cys Arg His Asn Pro Val Phe Gly Val Met Ser
      85      90

```

<210> 321
 <211> 98
 <212> PRT
 <213> Homo sapiens

<400> 321

```

Met Ser Val Ser Phe His Thr His Thr Lys Glu Leu Trp Thr Trp Met
1      5      10      15
Glu Asp Leu Gln Lys Glu Met Leu Glu Asp Val Cys Ala Asp Ser Val
      20      25      30
Asp Ala Val Gln Glu Leu Ile Lys Gln Phe Gln Gln Gln Gln Thr Ala
      35      40      45
Thr Leu Asp Ala Thr Leu Asn Val Ile Lys Glu Gly Glu Asp Leu Ile
      50      55      60
Gln Gln Leu Arg Ser Ala Pro Pro Ser Leu Gly Glu Pro Ser Glu Ala
65      70      75      80
Arg Ser Ala Trp Ala Glu Leu Ser Ser Gly Lys Cys Leu Gly Leu Asp
      85      90      95
Val Arg

```

<210> 322
 <211> 89
 <212> PRT
 <213> Homo sapiens

<400> 322

```

Met Val Ser Gly Asp Gly Phe Leu Val Ser Arg Pro Glu Ala Ile His
1      5      10      15
Leu Gly Pro Arg Gln Ala Val Arg Pro Ser Val Arg Ala Glu Ser Arg
      20      25      30
Arg Val Asp Gly Gly Gly Arg Ser Pro Arg Glu Pro Asp Gly Arg Gly
      35      40      45
Arg Ser Arg Gln Ala Arg Phe Ser Pro Tyr Pro Ile Pro Ala Val Glu
      50      55      60
Pro Asp Leu Leu Arg Ser Val Leu Gln Gln Arg Leu Ile Ala Leu Gly
65      70      75      80
Gly Val Ile Ala Ala Arg Ile Ser Val
      85

```

<210> 323
 <211> 123
 <212> PRT
 <213> Homo sapiens

<400> 323

186

Met Ala Leu Asn Asn Val Ser Leu Ser Ser Gly Asp Gln Arg Ser Arg
 1 5 10 15
 Val Ala Tyr Arg Ser Ser His Gly Asp Leu Arg Pro Arg Ala Ser Ala
 20 25 30
 Leu Ala Met Val Ser Gly Asp Gly Phe Leu Val Ser Arg Pro Glu Ala
 35 40 45
 Ile His Leu Gly Pro Arg Gln Ala Val Arg Pro Ser Val Arg Ala Glu
 50 55 60
 Ser Arg Arg Val Asp Gly Gly Arg Ser Pro Arg Glu Pro Asp Gly
 65 70 75 80
 Arg Gly Arg Ser Arg Gln Ala Arg Phe Ser Pro Tyr Pro Ile Pro Ala
 85 90 95
 Val Glu Pro Asp Leu Leu Arg Ser Val Leu Gln Gln Arg Leu Ile Ala
 100 105 110
 Leu Gly Gly Val Ile Ala Ala Arg Ile Ser Val
 115 120

<210> 324

<211> 172

<212> PRT

<213> Homo sapiens

<400> 324

Met Ala Gly Ala Ala Thr Gln Ala Ser Leu Glu Ser Ala Pro Arg Ile
 1 5 10 15
 Met Arg Leu Val Ala Glu Cys Ser Arg Ser Arg Ala Arg Ala Gly Glu
 20 25 30
 Leu Trp Leu Pro His Gly Thr Val Ala Thr Pro Val Phe Met Pro Val
 35 40 45
 Gly Thr Gln Ala Thr Met Lys Gly Ile Thr Thr Glu Gln Leu Asp Ala
 50 55 60
 Leu Gly Cys Arg Ile Cys Leu Gly Asn Thr Tyr His Leu Gly Leu Arg
 65 70 75 80
 Pro Gly Pro Glu Leu Ile Gln Lys Ala Asn Gly Leu His Gly Phe Met
 85 90 95
 Asn Trp Pro His Asn Leu Leu Thr Asp Ser Gly Gly Phe Gln Met Val
 100 105 110
 Ser Leu Val Ser Leu Ser Glu Val Thr Glu Glu Gly Val Arg Phe Arg
 115 120 125
 Ser Pro Tyr Asp Gly Asn Glu Thr Leu Leu Ser Pro Glu Lys Ser Val
 130 135 140
 Gln Ile Gln Asn Ala Leu Gly Ser Asp Ile Ile Met Gln Leu Asp Asp
 145 150 155 160
 Val Val Ser Ser Thr Val Thr Gly Pro Arg Val Glu
 165 170

<210> 325

<211> 20

<212> PRT

<213> Homo sapiens

<400> 325

Met Ala Ile Lys Phe Leu Glu Val Ile Lys Pro Phe Cys Val Ile Leu
 1 5 10 15
 Gln Leu Thr Ser
 20

<210> 326

<211> 23

<212> PRT

<213> Homo sapiens

<400> 326

187

Phe Leu Asn Asn Ser Ser Pro Gln Glu Pro Leu Lys Leu Pro Pro Gly
 1 5 10 15
 Gln Leu Ser Ser Thr His Arg
 20

<210> 327
 <211> 45
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 27
 <223> Xaa = Phe,Ser

<400> 327
 Met Glu Ser Asp Asp Ala Arg Ala Ala Arg Val Lys Glu Leu Leu Val
 1 5 10 15
 Asp Cys Tyr Lys Pro Thr Glu Ala Phe Ile Xaa Gly Leu Leu Asp Lys
 20 25 30
 Ile Arg Gly Met Gln Lys Leu Ser Thr Pro Gln Lys Lys
 35 40 45

<210> 328
 <211> 109
 <212> PRT
 <213> Homo sapiens

<400> 328
 Met Ala Ala Ser Ala Ala Arg Gly Ala Ala Ala Leu Arg Arg Ser Ile
 1 5 10 15
 Asn Gln Pro Val Ala Phe Val Arg Arg Ile Pro Trp Thr Ala Ala Ser
 20 25 30
 Ser Gln Leu Lys Glu His Phe Ala Gln Phe Gly His Val Arg Arg Cys
 35 40 45
 Ile Leu Pro Phe Asp Lys Glu Thr Gly Phe His Arg Gly Leu Gly Trp
 50 55 60
 Val Gln Phe Ser Ser Glu Glu Gly Leu Arg Asn Ala Leu Gln Gln Glu
 65 70 75 80
 Asn His Ile Ile Asp Gly Val Lys Val Gln Val His Thr Arg Arg Pro
 85 90 95
 Lys Leu Pro Gln Thr Ser Asp Asp Glu Lys Lys Asp Phe
 100 105

<210> 329
 <211> 51
 <212> PRT
 <213> Homo sapiens

<400> 329
 Met Ala Ala Ser Ala Ala Arg Gly Ala Ala Ala Leu His Lys Tyr Gln
 1 5 10 15
 Ser Ala Gly Cys Phe Cys Glu Lys Asn Ser Leu Asp Cys Gly Val Glu
 20 25 30
 Ser Ala Glu Arg Thr Leu Cys Thr Val Arg Pro Cys Gln Lys Val His
 35 40 45
 Phe Thr Phe
 50

<210> 330
 <211> 99
 <212> PRT
 <213> Homo sapiens

<400> 330

```

Pro Val Arg Gly Leu Asn Met Ser Glu Phe Val Cys Asn Leu Ser Ala
1           5           10           15
Arg Pro Tyr Val Tyr Asp Leu Ile Ala Val Ser Asn His Tyr Gly Ala
           20           25           30
Met Gly Val Gly His Tyr
           35

```

<210> 331

<211> 13

<212> PRT

<213> Homo sapiens

<400> 331

```

Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr
1           5           10

```

<210> 332

<211> 13

<212> PRT

<213> Homo sapiens

<400> 332

```

Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr
1           5           10

```

<210> 335

<211> 13

<212> PRT

<213> Homo sapiens

<400> 335

```

Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr
1           5           10

```

<210> 336

<211> 13

<212> PRT

<213> Homo sapiens

<400> 336

```

Met Leu Gly Phe Ala Lys Asn Trp Ile Phe Ser Ile Thr
1           5           10

```

<210> 337

<211> 109

<212> PRT

<213> Homo sapiens

<400> 337

```

Met Ala Ala Ser Ala Ala Arg Gly Ala Ala Ala Leu Arg Arg Ser Ile
1           5           10           15
Asn Gln Pro Val Ala Phe Val Arg Arg Xaa Pro Trp Thr Ala Ala Ser
           20           25           30
Ser Gln Leu Lys Glu His Phe Ala Gln Phe Gly His Val Arg Arg Cys
           35           40           45
Ile Leu Pro Phe Asp Lys Glu Thr Gly Phe His Arg Gly Leu Gly Trp
           50           55           60
Val Gln Phe Ser Ser Glu Glu Gly Leu Arg Asn Ala Leu Gln Gln Glu
65           70           75           80
Asn His Ile Ile Asp Gly Val Lys Val Gln Val His Thr Arg Arg Pro
           85           90           95

```

189

Lys Leu Pro Gln Thr Ser Asp Asp Glu Lys Lys Asp Phe
 100 105

<210> 338
 <211> 59
 <212> PRT
 <213> Homo sapiens

<400> 338
 Met Lys Phe Gly Cys Leu Ser Phe Arg Gln Pro Tyr Ala Gly Phe Val
 1 5 10 15
 Leu Asn Gly Ile Lys Thr Val Glu Thr Arg Trp Arg Pro Leu Leu Ser
 20 25 30
 Ser Gln Arg Asn Cys Thr Ile Ala Val His Ile Ala His Arg Asp Trp
 35 40 45
 Glu Gly Asp Ala Cys Arg Glu Leu Leu Val Glu
 50 55

<210> 339
 <211> 506
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 88..471

<220>
 <221> sig_peptide
 <222> 88..189
 <223> Von Heijne matrix
 score 13.6000003814697
 seq VLLLLSLLLTALA/SS

<400> 339
 acttatctgc agacttgttag gcagcaactc accctcactc agaggtcttc tggttctgga 60
 aacaactcta gctcagcctt ctccacc atg agc ctc aga ctt gat acc acc cct 114
 Met Ser Leu Arg Leu Asp Thr Thr Pro
 -30
 tcc tgt aac agt gcg aga cca ctt cat gcc ttg cag gtg ctg ctg ctt 162
 Ser Cys Asn Ser Ala Arg Pro Leu His Ala Leu Gln Val Leu Leu Leu
 -25 -20 -15 -10
 ctg tca ttg ctg ctg act gct ctg gct tcc tcc acc aaa gga caa act 210
 Leu Ser Leu Leu Leu Thr Ala Leu Ala Ser Ser Thr Lys Gly Gln Thr
 -5 1 5
 aag aga aac ttg gcg aaa ggc aaa gag gaa agt cta gac agt gac ttg 258
 Lys Arg Asn Leu Ala Lys Gly Lys Glu Glu Ser Leu Asp Ser Asp Leu
 10 15 20
 tat gct gaa ctc cgc tgc atg tgt ata aag aca acc tct gga att cat 306
 Tyr Ala Glu Leu Arg Cys Met Cys Ile Lys Thr Ser Gly Ile His
 25 30 35
 ccc aaa aac atc caa agt ttg gaa gtg atc ggg aaa gga acc cat tgc 354
 Pro Lys Asn Ile Gln Ser Leu Glu Val Ile Gly Lys Gly Thr His Cys
 40 45 50 55
 aac caa gtc gaa gtg ata gcc aca ctg aag gat ggg agg aaa atc tgc 402
 Asn Gln Val Glu Val Ile Ala Thr Leu Lys Asp Gly Arg Lys Ile Cys
 60 65 70
 ctg gac cca gat gct ccc aga atc aag aaa att gta cag aaa aaa ttg 450
 Leu Asp Pro Asp Ala Pro Arg Ile Lys Lys Ile Val Gln Lys Lys Leu
 75 80 85
 gca ggt gat gaa tct gct gat taatttgttc tgtttctgcc aaacttcttt 501
 Ala Gly Asp Glu Ser Ala Asp
 90

aactc

506

<210> 340
 <211> 523
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 41..235

<220>
 <221> sig_peptide
 <222> 41..94
 <223> Von Heijne matrix
 score 12.1999998092651
 seq LTLAVLFLTGSQA/RH

<400> 340
 ttagagactg cgagaaggag gtccccacg gcccttcagg atg aaa gct gcg gtg 55
 Met Lys Ala Ala Val
 -15
 ctg acc ttg gcc gtg ctc ttc ctg acg ggg agc cag gct cgg cat ttc 103
 Leu Thr Leu Ala Val Leu Phe Leu Thr Gly Ser Gln Ala Arg His Phe
 -10 -5 1
 tgg cag caa gat gaa ccc ccc cag agc ccc tgg gat cga gtg aag gac 151
 Trp Gln Gln Asp Glu Pro Pro Gln Ser Pro Trp Asp Arg Val Lys Asp
 5 10 15
 ctg gcc act gtg tac gtg gat gtg ctc aaa gac agc ggc aga gac tat 199
 Leu Ala Thr Val Tyr Val Asp Val Leu Lys Asp Ser Gly Arg Asp Tyr
 20 25 30 35
 gtg tcc cag ttt gaa ggc tcc gct tgg gaa aac agc taaacctaaa 245
 Val Ser Gln Phe Glu Gly Ser Ala Trp Glu Asn Ser
 40 45
 gctccttgac aactgggaca gcgtracctc crccttcagc aakctgcgcg aacagctcgg 305
 ccctgtgacc caggagttct gggataacct ggaaaaggag acagagggcc tgaggcagga 365
 gatgagcaag gatctggagg aggtgaaggc caasgtgcag ccctacctgg acgacttcca 425
 gaagaagtgg caggaggaga tggagctcta ccgccagaag gcagctttct taactatcct 485
 aacaagcctt ggaccaaag gaaataaagc tttttgat 523

<210> 341
 <211> 580
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 134..373

<220>
 <221> sig_peptide
 <222> 134..211
 <223> Von Heijne matrix
 score 12
 seq LLLALLLPTQIYS/SE

<400> 341
 atttgtggct ttcttggtat ataaggtctc gccggctcgc cgcgctcccc accttgctg 60
 cgcccgcccg gagccagcgg ttctccaagc acccagcctc ctgctagacg cgccgcgcac 120
 cgacggaggg gac atg ggc aga gca atg gtg gcc agg ctc ggg ctg ggg 169
 Met Gly Arg Ala Met Val Ala Arg Leu Gly Leu Gly
 -25 -20 -15
 ctg ctg ctg ctg gca ctg ctc cta ccc acg cag att tat tcc agt gaa 217

191

```

Leu Leu Leu Leu Ala Leu Leu Leu Pro Thr Gln Ile Tyr Ser Ser Glu
      -10      -5      1
aca aca act gga act tca agt aac tcc tcc cag agt act tcc aac tct      265
Thr Thr Thr Gly Thr Ser Ser Asn Ser Ser Gln Ser Thr Ser Asn Ser
      5      10      15
ggg ttg gcc cca aat cca act aat gcc acc acc aag gyg gct ggt ggt      313
Gly Leu Ala Pro Asn Pro Thr Asn Ala Thr Thr Lys Xaa Ala Gly Gly
      20      25      30
gcc ctg cag tca aca gcc agt ctc ttc gtg gtc tca ctc tct ctt ctg      361
Ala Leu Gln Ser Thr Ala Ser Leu Phe Val Val Ser Leu Ser Leu Leu
      35      40      45      50
cat ctc tac tct taagagactc aggccaagaa acgtcttcta aattttcccca      413
His Leu Tyr Ser
tcttctaaac ccaatccaaa tggcgtctgg aagtccaatg tggcaaggaa aaacaggtct      473
tcatcgaatc tactaattcc acacctttta ttgacacaga aaatgttgag aatcccaaat      533
ttgattgatt tgaagaacat gtgagaggkt tgactagatg atggatg      580

```

<210> 342

<211> 508

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 76..378

<220>

<221> sig_peptide

<222> 76..156

<223> Von Heijne matrix

score 11.8000001907349

seq LFLGLLLLPLVVA/FA

<400> 342

```

aattggccac agagacccag cccgagtttc ccatcgcaact gagcaactgag atcctgctgg      60
aagctctgcc gcagc atg agc tcc gca gcc ggg ttc tgc gcc tca cgc ccc      111
      Met Ser Ser Ala Ala Gly Phe Cys Ala Ser Arg Pro
      -25      -20
ggg ctg ctg ttc ctg ggg ttg ctg ctc ctg cca ctt gtg gtc gcc ttc      159
Gly Leu Leu Phe Leu Gly Leu Leu Leu Leu Pro Leu Val Val Ala Phe
-15      -10      -5      1
gcc agc gct gaa gct gaa gaa gat ggg gac ctg cag tgc ctg tgt gtg      207
Ala Ser Ala Glu Ala Glu Glu Asp Gly Asp Leu Gln Cys Leu Cys Val
      5      10      15
aag acc acc tcc cag gtc cgt ccc agg cac atc acc agc ctg gag gtg      255
Lys Thr Thr Ser Gln Val Arg Pro Arg His Ile Thr Ser Leu Glu Val
      20      25      30
atc aag gcc gga ccc cac tgc ccc act gcc caa ctg ata gcc acg ctg      303
Ile Lys Ala Gly Pro His Cys Pro Thr Ala Gln Leu Ile Ala Thr Leu
      35      40      45
aag aat gga agg aaa att tgc ttg gac ctg caa gcc ccg ctg tac aag      351
Lys Asn Gly Arg Lys Ile Cys Leu Asp Leu Gln Ala Pro Leu Tyr Lys
      50      55      60      65
aaa ata att aag aaa ctt ttg gag agt tagctactag ctgcctacgt      398
Lys Ile Ile Lys Lys Leu Leu Glu Ser
      70
gtgtgcattt gctatatagc atacttcttt tttccagttt caatctaact gtgaaagaac      458
ttctgatatt tgtgttatcc ttatgatttt aaataaacia aataaatcta      508

```

<210> 343

<211> 459

<212> DNA

<213> Homo sapiens

193

```

tat gat gcc aaa cca cag ggc cgg gat gac aga tac tgt gaa agc atc      318
Tyr Asp Ala Lys Pro Gln Gly Arg Asp Asp Arg Tyr Cys Glu Ser Ile
   15                               20                               25
atg agg aga cgg ggc ctg acc tca ccc tgc aaa gac atc aac aca ttt      366
Met Arg Arg Arg Gly Leu Thr Ser Pro Cys Lys Asp Ile Asn Thr Phe
   30                               35                               40                               45
att cat ggc aac aag cgc aga tca agg cca tct gtg aaa aca aga atg      414
Ile His Gly Asn Lys Arg Arg Ser Arg Pro Ser Val Lys Thr Arg Met
                   50                               55                               60
gaa acc ctc aca gag aaa acc taagaataag caagtcttct ttccaggtca      465
Glu Thr Leu Thr Glu Lys Thr
                   65
ccacttgcaa gctacatgga ggttcccttg gcctccatgc cagtaccgag ccacagcggg      525
gttcagaaac gttgttggtg cttgtgaa                                     553

```

<210> 345
 <211> 499
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 39..371

<220>
 <221> sig_peptide
 <222> 39..86
 <223> Von Heijne matrix
 score 11.5
 seq ILLSVALLAFSSA/QD

```

<400> 345
agcataaagt tgggagtgac accagagcct tctgcaag atg ctt ctg att ctg ctg      56
                               Met Leu Leu Ile Leu Leu
                               -15
tca gtg gcc ctg ctg gcc ttc agc tca gct cag gac tta gat gaa gat      104
Ser Val Ala Leu Leu Ala Phe Ser Ser Ala Gln Asp Leu Asp Glu Asp
-10                               -5                               1                               5
gtc agc caa gaa gac gtt ccc ttg gta ata tca gat gga gga gac tct      152
Val Ser Gln Glu Asp Val Pro Leu Val Ile Ser Asp Gly Gly Asp Ser
                   10                               15                               20
gag cag ttc ata gat gag gag cgt cag gga cca cct ttg gga gga cag      200
Glu Gln Phe Ile Asp Glu Glu Arg Gln Gly Pro Pro Leu Gly Gly Gln
                   25                               30                               35
caa tct caa ccc tct gct ggt gat ggg aac cag rat gat ggc cct cag      248
Gln Ser Gln Pro Ser Ala Gly Asp Gly Asn Gln Xaa Asp Gly Pro Gln
                   40                               45                               50
cag gga cca ccc caa caa gga ggc cag cag caa caa ggt cca cca cct      296
Gln Gly Pro Pro Gln Gln Gly Gly Gln Gln Gln Gln Gly Pro Pro Pro
                   55                               60                               65                               70
cct cag gga aag cca car gga cca cct ccc caa ggg ggc cgc cca caa      344
Pro Gln Gly Lys Pro Gln Gly Pro Pro Pro Gln Gly Gly Arg Pro Gln
                   75                               80                               85
gga cct cca cag ggg cag tct cct cag taatctagga ttcaatgaca      391
Gly Pro Pro Gln Gly Gln Ser Pro Gln
                   90                               95
ggaagtgaat aagaagatga cagtgtttca aatgccttga aacataatgt gatcatgctc      451
taacttcaat ataccaataa aataatcagc ttgcaaaaaa aaaaaaaa                                     499

```

<210> 346
 <211> 398
 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS

<222> 36..353

<220>

<221> sig_peptide

<222> 36..83

<223> Von Heijne matrix

score 11.5

seq ILLSVALLAFSSA/QD

<400> 346

```

ataaagttagg gagtgacacc agagccttct gcaag atg ctt ctg att ctg ctg      53
                               Met Leu Leu Ile Leu Leu
                               -15
tca gtg gcc ctg ctg gcc ttc agc tca gct cag gac tta gat gaa gat      101
Ser Val Ala Leu Leu Ala Phe Ser Ser Ala Gln Asp Leu Asp Glu Asp
-10                               -5                               5
gtc agc caa gaa gac gtt ccc ttg gta ata tca gat gga gga gac tct      149
Val Ser Gln Glu Asp Val Pro Leu Val Ile Ser Asp Gly Gly Asp Ser
          10          15          20
gag cag ttc ata gat gag gag cgt cag gga cca cct ttg gga gga cag      197
Glu Gln Phe Ile Asp Glu Glu Arg Gln Gly Pro Pro Leu Gly Gly Gln
          25          30          35
caa tct caa ccc tct gct ggt gat ggg aac cag gat gat ggc cct cag      245
Gln Ser Gln Pro Ser Ala Gly Asp Gly Asn Gln Asp Asp Gly Pro Gln
          40          45          50
cag gga cca ccc caa caa gga ggc cag cag caa caa ggt cca cca cct      293
Gln Gly Pro Pro Gln Gln Gly Gly Gln Gln Gln Gln Gly Pro Pro Pro
55          60          65          70
cct cag gga aag cca caa gga cca ccc caa cag gga ggc cag tct tgt      341
Pro Gln Gly Lys Pro Gln Gly Pro Pro Gln Gln Gly Gly Gln Ser Cys
          75          80          85
tgc tgt gac aag taacagcctt ttttttaaatt tgttttttat acaaaaaaaaa aaaaa      398
Cys Cys Asp Lys
          90

```

<210> 347

<211> 525

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 48..470

<220>

<221> sig_peptide

<222> 48..119

<223> Von Heijne matrix

score 10.89999996185303

seq LAGLCCLVPVSLA/ED

<400> 347

```

ctcctcagct tcaggcacca ccactgacct gggacagtga atcgaca atg ccg tct      56
                               Met Pro Ser
tct gtc tcg tgg ggc atc ctc ctg ctg gca ggc ctg tgc tgc ctg gtc      104
Ser Val Ser Trp Gly Ile Leu Leu Leu Ala Gly Leu Cys Cys Leu Val
-20          -15          -10
cct gtc tcc ctg gct gag gat ccc cag gga gat gct gcc cag aag aca      152
Pro Val Ser Leu Ala Glu Asp Pro Gln Gly Asp Ala Ala Gln Lys Thr
-5          1          5          10

```

195

```

gat aca tcc cac cat gat cag gat cac cca acc ttc aac aag atc acc      200
Asp Thr Ser His His Asp Gln Asp His Pro Thr Phe Asn Lys Ile Thr
      15                                20                                25
ccc aac ctg gct gag ttc gcc ttc agc cta tac cgc cag ctg gca cac      248
Pro Asn Leu Ala Glu Phe Ala Phe Ser Leu Tyr Arg Gln Leu Ala His
      30                                35                                40
cag tcc aac agc acc aat atc ttc ttc tcc cca gtg agc atc gct aca      296
Gln Ser Asn Ser Thr Asn Ile Phe Phe Ser Pro Val Ser Ile Ala Thr
      45                                50                                55
gcc ttt gca atg ctc tcc ctg ggg acc aag gct gac act cac gat gaa      344
Ala Phe Ala Met Leu Ser Leu Gly Thr Lys Ala Asp Thr His Asp Glu
      60                                65                                70                                75
atc ctg gag ggc ctg aat ttc aac ctc acg gag att ccg gag gct cag      392
Ile Leu Glu Gly Leu Asn Phe Asn Leu Thr Glu Ile Pro Glu Ala Gln
      80                                85                                90
atc cat gaa ggc ttc cag gaa ctc ctc cgt acc ctc aac cag cca gac      440
Ile His Glu Gly Phe Gln Glu Leu Leu Arg Thr Leu Asn Gln Pro Asp
      95                                100                                105
agc cag ctc cag ctg acc acc ggc aag aat taatgctttc ttaactaagt      490
Ser Gln Leu Gln Leu Thr Thr Gly Lys Asn
      110                                115
ttactttgag tattaacaag tagcttatac tgagg      525

```

<210> 348

<211> 500

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 91..462

<220>

<221> sig_peptide

<222> 91..180

```

<223> Von Heijne matrix
      score 9.89999961853027
      seq LLFGFTLVSGTGA/EK

```

<400> 348

```

accctaccc cagccccct cccgcgcgcg cggttaaatc cccgcacctg agcatcggt      60
cacacctgca cccgccccgg gcatagcacc atg cct gct tgt cgc cta ggc ccg      114
                                Met Pro Ala Cys Arg Leu Gly Pro
                                -30                                -25
cta gcc gcc gcc ctc ctc ctc agc ctg ctg ctg ttc ggc ttc acc cta      162
Leu Ala Ala Ala Leu Leu Leu Ser Leu Leu Leu Phe Gly Phe Thr Leu
      -20                                -15                                -10
gtc tca ggc aca gga gca gag aag act ggc gtg tgc ccc gag ctc cag      210
Val Ser Gly Thr Gly Ala Glu Lys Thr Gly Val Cys Pro Glu Leu Gln
      -5                                1                                5                                10
gct gac cag aac tgc acg caa gag tgc gtc tcg gac agc gaa tgc gcc      258
Ala Asp Gln Asn Cys Thr Gln Glu Cys Val Ser Asp Ser Glu Cys Ala
      15                                20                                25
gac aac ctc aag tgc tgc agc gcg ggc tgt ggc acc ttc tgc tct ctg      306
Asp Asn Leu Lys Cys Cys Ser Ala Gly Cys Ala Thr Phe Cys Ser Leu
      30                                35                                40
ccc aat gat aag gag ggt tcc tgc ccc cag gtg aac att aac ttt ccc      354
Pro Asn Asp Lys Glu Gly Ser Cys Pro Gln Val Asn Ile Asn Phe Pro
      45                                50                                55
cag ctc ggc ctc tgt cgg gac cag tgc cag gtg gac agc cag tgt cct      402
Gln Leu Gly Leu Cys Arg Asp Gln Cys Gln Val Asp Ser Gln Cys Pro
      60                                65                                70
ggc cag atg aaa tgc tgc cgc aat ggc tgt ggg aag gtg tcc tgt gtc      450

```

196

Gly Gln Met Lys Cys Cys Arg Asn Gly Cys Gly Lys Val Ser Cys Val
 75 80 85 90
 act ccc aat ttc tgagctccag ccaccaccag gctgagcagt gaggagag 500
 Thr Pro Asn Phe

<210> 349
 <211> 519
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 28..399

<220>
 <221> sig_peptide
 <222> 28..117
 <223> Von Heijne matrix
 score 9.89999961853027
 seq LLFGFTLVSGTGA/EK

<400> 349
 acctgcaccc cgcccgggca tagcacc atg cct gct tgt cgc cta ggc ccg cta 54
 Met Pro Ala Cys Arg Leu Gly Pro Leu
 -30 -25
 gcc gcc gcc ctc ctc ctc agc ctg ctg ctg ttc ggc ttc acc cta gtc 102
 Ala Ala Ala Leu Leu Leu Ser Leu Leu Leu Phe Gly Phe Thr Leu Val
 -20 -15 -10
 tca ggc aca gga gca gag aag act ggc gtg tgc ccc gag ctc cag gct 150
 Ser Gly Thr Gly Ala Glu Lys Thr Gly Val Cys Pro Glu Leu Gln Ala
 -5 1 5 10
 gac cag aac tgc acg caa gag tgc gtc tcg gac agc gaa tgc gcc gac 198
 Asp Gln Asn Cys Thr Gln Glu Cys Val Ser Asp Ser Glu Cys Ala Asp
 15 20 25
 aac atc aag tgc tgc agc gcg ggc tgt gcc acc ttc tgc tct ctg ccc 246
 Asn Ile Lys Cys Cys Ser Ala Gly Cys Ala Thr Phe Cys Ser Leu Pro
 30 35 40
 aat gat aag gag ggt tcc tgc ccc cag gtg aac att aac ttt ccc cag 294
 Asn Asp Lys Glu Gly Ser Cys Pro Gln Val Asn Ile Asn Phe Pro Gln
 45 50 55
 ctc ggc ctc tgt cgg gac cag tgc cag gtg gac agc cag tgt cct ggc 342
 Leu Gly Leu Cys Arg Asp Gln Cys Gln Val Asp Ser Gln Cys Pro Gly
 60 65 70 75
 cag atg aaa tgc tgc cgc aat ggc tgt ggg aag gtg tcc tgt gtc act 390
 Gln Met Lys Cys Cys Arg Asn Gly Cys Gly Lys Val Ser Cys Val Thr
 80 85 90
 ccc aat ttc tgagctccag ccacaggctg agcagtgagg agagaaagtt 439
 Pro Asn Phe
 tctgcctggc cctgcattctg gttccagccc acctgcctc ccctttttctg gggactctgt 499
 attccctctt gggctgacca 519

<210> 350
 <211> 482
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 40..246

<220>
 <221> sig_peptide
 <222> 40..90

<223> Von Heijne matrix
score 9.5
seq FVFLLFVISLAAA/AH

<400> 350

```

aaaaagtagg acaccggagc tagcagtggc aaccactt atg ttt cct ggc ttt      54
                                   Met Phe Pro Gly Phe
                                   -15
gtt ttt ttg ctc ttt gta ata agt ctt gct gct gct gct cat ctt tgg    102
Val Phe Leu Leu Phe Val Ile Ser Leu Ala Ala Ala His Leu Trp
      -10                    -5                    1
gtc ctg gcc gcc ttt atg ggc cgt atc acc gtg aag gtc tgc agc ttc    150
Val Leu Ala Ala Phe Met Gly Arg Ile Thr Val Lys Val Cys Ser Phe
5          10          15          20
act cct gag gcc agc aag acc gtg agc cca cca gaa gga gcg aac aac    198
Thr Pro Glu Ala Ser Lys Thr Val Ser Pro Pro Glu Gly Ala Asn Asn
      25          30          35
tcg aga cgc act gct ttt aag agc tgt aac act cac cac aaa ggt ctg    246
Ser Arg Arg Thr Ala Phe Lys Ser Cys Asn Thr His His Lys Gly Leu
      40          45          50
tagcatcact cctgaagtca gcgggaccac gaacctacca gaaagaagaa actccatata    306
gatctgaaaa tctgaaggaa caaactccgg acacaccatc tttaagaact gtaacactca    366
ctgcgagggt ccacggcttc atttttgaag tcagccagac caagaacca ccaattccgc    426
acacacttct gctgtttcgt tccattctca aggaaaactg tgatgggtcc attttt    482

```

<210> 351
<211> 462
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 114..344

<220>
<221> sig_peptide
<222> 114..212
<223> Von Heijne matrix
score 9.30000019073486
seq MVLLSFALTLCSA/FW

<400> 351

```

wctgctcact gctgggtact gttctgctgt gggtgcccag gaagggacta cacctcttcg    60
cagtgtttta tacctttggt aatatcgcat caattgggag taccatcttc ctc atg    116
                                   Met
gga cca gtg aaa cag ctg aag cga atg ttt gag cct act cgt ttg att    164
Gly Pro Val Lys Gln Leu Lys Arg Met Phe Glu Pro Thr Arg Leu Ile
      -30          -25          -20
gca act atc atg gtg ctg ttg agt ttt gca ctt acc ctg tgt tct gcc    212
Ala Thr Ile Met Val Leu Leu Ser Phe Ala Leu Thr Leu Cys Ser Ala
      -15          -10          -5
ttt tgg tgg cat aac atg gga ctt gca ctt atc ttc tgc att ttg cag    260
Phe Trp Trp His Asn Met Gly Leu Ala Leu Ile Phe Cys Ile Leu Gln
1          5          10          15
tct ttg gca ttg acg tgg tac agc ctt tcc ttc ata cca ttt gca agg    308
Ser Leu Ala Leu Thr Trp Tyr Ser Leu Ser Phe Ile Pro Phe Ala Arg
      20          25          30
gat gct gtg aag aag tgt ttt gcc gtg tgt ctt gca taattcatgg    354
Asp Ala Val Lys Lys Cys Phe Ala Val Cys Leu Ala
      35          40
ccagttttat gaagcttttg aaggcactat ggacagaagc tgggtggacag ttttgtaact    414
atcttcgaaa cctctgtctt acagacatgt gccttttatc ttgcagca    462

```

<210> 352
 <211> 496
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 29..466

<220>
 <221> sig_peptide
 <222> 29..106
 <223> Von Heijne matrix
 score 8.89999961853027
 seq GLLWAFCAPGARA/EE

<400> 352
 cgtctgcttc ggagaccgta agatattg atg acc atg aga tcc ctg ctc aga 52
 Met Thr Met Arg Ser Leu Leu Arg
 -25 -20
 acc ccc ttc ctg tgt ggc ctg ctc tgg gcc ttt tgt gcc cca ggc gcc 100
 Thr Pro Phe Leu Cys Gly Leu Leu Trp Ala Phe Cys Ala Pro Gly Ala
 -15 -10 -5
 agg gct gag gag cct gca gcc agc ttc tcc caa ccc ggc agc atg ggc 148
 Arg Ala Glu Glu Pro Ala Ala Ser Phe Ser Gln Pro Gly Ser Met Gly
 1 5 10
 ctg gat aag aac aca gtg cac gac caa gag cat atc atg gag cat cta 196
 Leu Asp Lys Asn Thr Val His Asp Gln Glu His Ile Met Glu His Leu
 15 20 25 30
 gaa ggt gtc atc aac aaa cca gag gcg gag atg tcg cca caa gaa ttg 244
 Glu Gly Val Ile Asn Lys Pro Glu Ala Glu Met Ser Pro Gln Glu Leu
 35 40 45
 cag ctc cat tac ttc aaa atg cat gat tat gat ggc aat aat ttg ctt 292
 Gln Leu His Tyr Phe Lys Met His Asp Tyr Asp Gly Asn Asn Leu Leu
 50 55 60
 gat ggc tta gaa ctc tcc aca gcc atc act cat gtc cat aag gag gaa 340
 Asp Gly Leu Glu Leu Ser Thr Ala Ile Thr His Val His Lys Glu Glu
 65 70 75
 ggg agt gaa cag gca cca cta atg agt gaa gat gaa ctg att aac ata 388
 Gly Ser Glu Gln Ala Pro Leu Met Ser Glu Asp Glu Leu Ile Asn Ile
 80 85 90
 ata gat ggt gtt ttg aga gat gat gac aag aac aat gat gga tac att 436
 Ile Asp Gly Val Leu Arg Asp Asp Asp Lys Asn Asn Asp Gly Tyr Ile
 95 100 105 110
 gac tat gct gaa ttt gca aaa tca ctg cag tagatgttat ttggcatctc 486
 Asp Tyr Ala Glu Phe Ala Lys Ser Leu Gln
 115 120
 ctggttatat 496

<210> 353
 <211> 499
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 155..340

<220>
 <221> sig_peptide
 <222> 155..292
 <223> Von Heijne matrix
 score 8.60000038146973

seq AVLLLLILFAIVFG/LL

<400> 353

```

cttttctctt caacagttgc ttctttgagt caggggtgcag ctctggtcac ctggcggcct      60
cttcagctca gccctccaca aagtgtgagc ctgaaggacc accctgaatt gcccttgtag      120
gaccagaac agcwaccagc agaatcagat tctc atg gac caa ytg gta ttc aaa      175
                               Met Asp Gln Leu Val Phe Lys
                               -45                               -40

gag aca atc tgg aat gat gcg ttc tgg cag aac ccc tgg gac cag ggg      223
Glu Thr Ile Trp Asn Asp Ala Phe Trp Gln Asn Pro Trp Asp Gln Gly
                               -35                               -30                               -25

ggc ctg gca gtg att atc tta ttc atc acc gct gtc ctg ctt ctc atc      271
Gly Leu Ala Val Ile Ile Leu Phe Ile Thr Ala Val Leu Leu Leu Ile
                               -20                               -15                               -10

tta ttt gcc atc gtg ttt ggt tta ctc act tcc aca gaa aac act cag      319
Leu Phe Ala Ile Val Phe Gly Leu Leu Thr Ser Thr Glu Asn Thr Gln
                               -5                               1                               5

tgt gaa gcg ggt gaa gag gag tgacctgact tgctggggac tgagatggca      370
Cys Glu Ala Gly Glu Glu
10                               15

gcaggggagg cgagctgacc tgcccccatt ccagtggtgg gcccttcgag gttccctctg      430
gctcaggggc caagcctggt gtcttccttt cccaccagga aaaagtctag taaaatactg      490
tatctggct      499

```

<210> 354

<211> 537

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 13..498

<220>

<221> sig_peptide

<222> 13..153

<223> Von Heijne matrix

score 8.30000019073486

seq LALSSLLSLLLFA/GM

<400> 354

```

tagcgctgac gc atg cgc ata gct aac cgc acc cgg ttc agc tcg cct ttc      51
                               Met Arg Ile Ala Asn Arg Thr Arg Phe Ser Ser Pro Phe
                               -45                               -40                               -35

ttg gcc aga ggc gcc ggt tgg act cac ggg cgg ggc atg atg gtg gtg      99
Leu Ala Arg Gly Ala Gly Trp Thr His Gly Arg Gly Met Met Val Val
                               -30                               -25                               -20

ggt acg ggc acc tcg ctg gcg ctc tcc tcc ctc ctg tcc ctg ctg ctc      147
Gly Thr Gly Thr Ser Leu Ala Leu Ser Ser Leu Leu Ser Leu Leu Leu
                               -15                               -10                               -5

ttt gct ggg atg cag atg tac agc cgt cag ctg gcc tcc acc gag tgg      195
Phe Ala Gly Met Gln Met Tyr Ser Arg Gln Leu Ala Ser Thr Glu Trp
1                               5                               10

ctc acc atc cag ggc gcc ctg ctt ggt tcg ggt ctc ttc gtg ttc tcg      243
Leu Thr Ile Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe Val Phe Ser
15                               20                               25                               30

ctc act gcc ttc aat aat ctg gag aat ctt gtc ttt ggc aaa gga ttc      291
Leu Thr Ala Phe Asn Asn Leu Glu Asn Leu Val Phe Gly Lys Gly Phe
                               35                               40                               45

caa gca aag atc ttc cct gag att ctc ctg tgc ctc ctg ttg gct ctc      339
Gln Ala Lys Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu Leu Ala Leu
                               50                               55                               60

ttt gca tct ggc ctc atc cac cga gtc tgt gtc acc acc tgc ttc atc      387

```


200

```

Phe Ala Ser Gly Leu Ile His Arg Val Cys Val Thr Thr Cys Phe Ile
      65                      70                      75
ttc tcc atg gtt ggt ctg tac tac atc aac aag atc tcc tcc acc ctg      435
Phe Ser Met Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser Ser Thr Leu
      80                      85                      90
tac cag gca gca gct cca gtc ctc aca cca gcc aag gtc aca ggc aag      483
Tyr Gln Ala Ala Ala Pro Val Leu Thr Pro Ala Lys Val Thr Gly Lys
95                      100                      105                      110
agc aag aag aga aac tgaccctgaa tgttcaataa agttgattct ttgcaaaaa      537
Ser Lys Lys Arg Asn
                      115

```

<210> 355
 <211> 514
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 132..401

<220>
 <221> sig_peptide
 <222> 132..236
 <223> Von Heijne matrix
 score 8.30000019073486
 seq LGLGLVTLTASWA/AL

```

<400> 355
ccttcagckc ccggccgaga gcccggtggc tgcgactgak agcctgggta ctgccggcac      60
ctttggatcc tcggccaatt ctctgcttcg cctacacttg gcgtgcgggtg taccagcctc      120
cccgcgtcac c atg gaa aca gga gcc tct gca tcc atc cca gag ctg atc      170
          Met Glu Thr Gly Ala Ser Ala Ser Ile Pro Glu Leu Ile
          -35                      -30                      -25
tgt gaa gct atg aga aga atc tgg agc ctg gga ttg ggg ttg gtg act      218
Cys Glu Ala Met Arg Arg Ile Trp Ser Leu Gly Leu Gly Leu Val Thr
          -20                      -15                      -10
ctg acg gcc agc tgg gca gct ctt ttc cac gat ggc ttt gcg gtt ctt      266
Leu Thr Ala Ser Trp Ala Ala Leu Phe His Asp Gly Phe Ala Val Leu
          -5                      1                      5                      10
gga gga aac att gtg agc gat ctc agc aca gta aga ttt gtt gca caa      314
Gly Gly Asn Ile Val Ser Asp Leu Ser Thr Val Arg Phe Val Ala Gln
          15                      20                      25
cag cag cac ttc cag ctc ctt gac gtg tgg acc agg aat ttc cgg aag      362
Gln Gln His Phe Gln Leu Leu Asp Val Trp Thr Arg Asn Phe Arg Lys
          30                      35                      40
cca ctg ggc agc atg tgc ttt gtt ttc ttg ttg ctc cca taatcaatgt      411
Pro Leu Gly Ser Met Cys Phe Val Phe Leu Leu Leu Pro
          45                      50                      55
tgggcatcaa gatctaacc ttgaacctc cacgaaccct gttgtcaata cctctgggtt      471
tccgccggtt acgcttaatt ttgacatatc ggtcagactg gtg      514

```

<210> 356
 <211> 589
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 41..223

<220>
 <221> sig_peptide

<222> 41..115

<223> Von Heijne matrix

score 8.10000038146973

seq ATLLLVLCCLQLGA/NK

<400> 356

```

accacttctc tgggacacat tgcctttctgt tttctccagc atg cgc ttg ctc cag      55
                                         Met Arg Leu Leu Gln
                                         -25
ctc ctg ttc agg gcc agc cct gcc acc ctg ctc ctg gtt ctc tgc ctg      103
Leu Leu Phe Arg Ala Ser Pro Ala Thr Leu Leu Leu Val Leu Cys Leu
-20                               -15                -10                -5
cag ttg ggg gcc aac aaa gct cag gac aac act cgg aag atc ata ata      151
Gln Leu Gly Ala Asn Lys Ala Gln Asp Asn Thr Arg Lys Ile Ile Ile
                               1                   5                   10
aag aat ttt gac att ccc aag tca gta cgt cca aat gac gaa gya ctg      199
Lys Asn Phe Asp Ile Pro Lys Ser Val Arg Pro Asn Asp Glu Xaa Leu
                               15                20                25
cag tgc ttg cag ttc aaa cag aat tgaaagaatg catggtgggtt aaaacttacc      253
Gln Cys Leu Gln Phe Lys Gln Asn
                               30                35
tcattagcag catccctcta caaggtgcat ttaactataa gtatactgcc tgcctatgtg      313
acgacaatcc aaaaaccttc tactgggact tttacaccaa cagaactgtg caaattgcag      373
ccgtcgttga tgttattcgg gaattaggca tctgcctgat gatgctgctg taatcccat      433
caaaaacaac cggttttata ctattgaaat cctaaaggta gaataatgga agccctgtct      493
gtttgccaca cccaggtgat ttcctctaaa gaaacttggc tggaatttct gctgtgggtct      553
ataaaataaa cttcttaaca tacaaaaaaa aaaaaa      589

```

<210> 357

<211> 465

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 19..378

<220>

<221> sig_peptide

<222> 19..114

<223> Von Heijne matrix

score 7.59999990463257

seq VTLGIGFFALASA/LW

<400> 357

```

cctaatacmw aaagtcacat atg gct tca cct aaa ggt ttt ttt aat tac ttg      51
                               Met Ala Ser Pro Lys Gly Phe Phe Asn Tyr Leu
                               -30                               -25
acc tat ttt ctt gct gct ggt gct gtc act ttg gga att ggt ttc ttt      99
Thr Tyr Phe Leu Ala Ala Gly Ala Val Thr Leu Gly Ile Gly Phe Phe
-20                               -15                -10
gct ttg gca tca gct ttg tgg ttc ctg att tgc aaa cga aga gaa ata      147
Ala Leu Ala Ser Ala Leu Trp Phe Leu Ile Cys Lys Arg Arg Glu Ile
-5                               1                   5                   10
ttt caa aat tcc aaa ttt aaa gca att gat gag aga tgc agg caa aga      195
Phe Gln Asn Ser Lys Phe Lys Ala Ile Asp Glu Arg Cys Arg Gln Arg
                               15                20                25
cca tca atg gcg aaa att aaa tct cat tct cag tgt gtt ttt att tct      243
Pro Ser Met Ala Lys Ile Lys Ser His Ser Gln Cys Val Phe Ile Ser
                               30                35                40
cga aat ttt cat act ggg aga ttc caa tta cag caa tta aag atc att      291
Arg Asn Phe His Thr Gly Arg Phe Gln Leu Gln Gln Leu Lys Ile Ile
                               45                50                55

```

202

```

cta aag atg aac ccc aac ttg caa caa aaa ata tca ttt gtg atc cct      339
Leu Lys Met Asn Pro Asn Leu Gln Gln Lys Ile Ser Phe Val Ile Pro
60                               65                               70                               75
cag aga cca gct cca caa caa atc gca gca gtg tta cat taagcttattc      388
Gln Arg Pro Ala Pro Gln Gln Ile Ala Ala Val Leu His
                               80                               85
aacattacca tctgattctt attacagcca aagtatagaa gcagctgatg actgggttttc      448
tgatgattct ctagtga                                              465

```

<210> 358
 <211> 438
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 84..317

<220>
 <221> sig_peptide
 <222> 84..140
 <223> Von Heijne matrix
 score 7.59999990463257
 seq ALVLALMISMISA/DS

```

<400> 358
atagaaaagg acatctcttg agacttcact tcagcttcac tgacttcttg actctctctt      60
tgagtaaaag gactcagcca act atg aag ttt ttt gtc ttt gct tta gtc ttg      113
                               Met Lys Phe Phe Val Phe Ala Leu Val Leu
                               -15                               -10
gct ctc atg att tcc atg att agc gct gat tca cat gaa aag aga cat      161
Ala Leu Met Ile Ser Met Ile Ser Ala Asp Ser His Glu Lys Arg His
                               -5                               1                               5
cat ggg tat aga aga aaa ttc cat gaa aag cat cat tca tac cat atc      209
His Gly Tyr Arg Arg Lys Phe His Glu Lys His His Ser Tyr His Ile
                               10                               15                               20
aca cta cta cca ctt ttt gaa gaa tca tca aag agc aat gca aat gaa      257
Thr Leu Leu Pro Leu Phe Glu Glu Ser Ser Lys Ser Asn Ala Asn Glu
                               25                               30                               35
aaa cac tat aat tta ctg tat act ctt tgt ttc agg ata ctt gcc ttt      305
Lys His Tyr Asn Leu Leu Tyr Thr Leu Cys Phe Arg Ile Leu Ala Phe
40                               45                               50                               55
tca att gtc act tgatgatata attgcaattt aaactgttaa gctgtgttca      357
Ser Ile Val Thr
gtactgtttc tgaataatag aaatcacttc tctaaaagca ataaatttca agcacatttt      417
taaataaaaa aaatmtmaaa a                                              438

```

<210> 359
 <211> 537
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 57..317

<220>
 <221> sig_peptide
 <222> 57..116
 <223> Von Heijne matrix
 score 7.40000009536743
 seq LLVFLAGFPVLDA/ND

<400> 359

gatttctccc ggaacctctg ctcagcctgg tgaaccacac aggccagcgc tctgac atg 59
Met
-20

cag aag gtg acc ctg ggc ctg ctt gtg ttc ctg gca ggc ttt cct gtc 107
Gln Lys Val Thr Leu Gly Leu Leu Val Phe Leu Ala Gly Phe Pro Val
-15 -10 -5

ctg gac gcc aat gac cta gaa gat aaa aac agt cct ttc tac tat gac 155
Leu Asp Ala Asn Asp Leu Glu Asp Lys Asn Ser Pro Phe Tyr Tyr Asp
1 5 10

tgg cac agc ctc cag gtt ggc ggg ctc atc tgc gct ggg gtt ctg tgc 203
Trp His Ser Leu Gln Val Gly Gly Leu Ile Cys Ala Gly Val Leu Cys
15 20 25

gcc atg ggc atc atc atc gtc atg agt gca aaa tgc aaa tgc aag ttt 251
Ala Met Gly Ile Ile Ile Val Met Ser Ala Lys Cys Lys Cys Lys Phe
30 35 40 45

ggc cag aag tcc ggt cac cat cca ggg gag act cca cct ctc atc acc 299
Gly Gln Lys Ser Gly His His Pro Gly Glu Thr Pro Pro Leu Ile Thr
50 55 60

cca ggc tca gcc caa agc tgatgaggac agaccagctg aaattgggtg 347
Pro Gly Ser Ala Gln Ser
65

gaggaccgtt ctctgtcccc aggtcctgtc tctgcacaga aacttgaact ccaggatgga 407
attcttctctc ctctgtctggg actcctttgc atggcagggc tcatctcacc tctcgcaaga 467
gggtctctttt gttcaatttt ttttaaatcta aaatgattgt gcctctgccc aaaaaaaaaa 527
aaaaaaaaagct 537

<210> 360

<211> 499

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 69..383

<220>

<221> sig_peptide

<222> 69..125

<223> Von Heijne matrix
score 7.19999980926514
seq VSIMLLLVTVSDC/AV

<400> 360

acaaggctga gcgggaggaa gcgagaggca tctaagcagg cagtgttttg ccttcacccc 60
aagtgacc atg aga ggt gcc acg cga gtc tca atc atg ctc ctc cta gta 110
Met Arg Gly Ala Thr Arg Val Ser Ile Met Leu Leu Leu Val
-15 -10

act gtg tct gac tgt gct gtg atc aca ggg gcc tgt gag cga gat gtc 158
Thr Val Ser Asp Cys Ala Val Ile Thr Gly Ala Cys Glu Arg Asp Val
-5 1 5 10

cag tgt ggg gca ggc acc tgc tgt gcc atc agc ctg tgg ctt cga ggg 206
Gln Cys Gly Ala Gly Thr Cys Cys Ala Ile Ser Leu Trp Leu Arg Gly
15 20 25

ctg cgg atg tgc acc ccg ctg ggg cgg waa ggc gag gag tgc cac ccc 254
Leu Arg Met Cys Thr Pro Leu Gly Arg Xaa Gly Glu Glu Cys His Pro
30 35 40

ggc agc cac aag atc ccc ttc ttc agg aaa cgc aag cac cac acc tgt 302
Gly Ser His Lys Ile Pro Phe Phe Arg Lys Arg Lys His His Thr Cys
45 50 55

cct tgc ttg ccc aac ctg ctg tgc tcc agg ttc ccg gac ggc agg tac 350
Pro Cys Leu Pro Asn Leu Leu Cys Ser Arg Phe Pro Asp Gly Arg Tyr
60 65 70 75

204

cgc tgc tcc atg gac ttg aag aac atc aat ttt taggcgcttg cctgggtctca 403
 Arg Cys Ser Met Asp Leu Lys Asn Ile Asn Phe
 80 85
 ggataccac catccttttc ctgagcacag cctggatttt tatttctgcc atgaaaccca 463
 gctcccatga ctctccaggt ccctasactg actacc 499

<210> 361
 <211> 479
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 129..434

<220>
 <221> sig_peptide
 <222> 129..236
 <223> Von Heijne matrix
 score 7
 seq PFFLITLVGVVVA/VV

<400> 361
 tccggggccc gggagccaac cgagggcggtt cctgtcgggg ctgcagcggc gggagggagc 60
 ccagtggagg cgccctcccg aagcgccact gcycatgctg accacccagc cctccggctg 120
 ctgatgtc atg agt aac acc act gtg ccc aat gcc ccc cag gcc aac agc 170
 Met Ser Asn Thr Thr Val Pro Asn Ala Pro Gln Ala Asn Ser
 -35 -30 -25
 gac tcc atg gtg ggc tat gtg ttg ggg ccc ttc ttc ctc atc acc ctg 218
 Asp Ser Met Val Gly Tyr Val Leu Gly Pro Phe Phe Leu Ile Thr Leu
 -20 -15 -10
 gtc ggg gtg gtg gtg gct gtg gta atg tat gta cag aag aaa aag cgg 266
 Val Gly Val Val Val Ala Val Val Met Tyr Val Gln Lys Lys Lys Arg
 -5 1 5 10
 gtg gac cgg ctg cgc cat cac ctg ctc ccc atg tac agc tat gac cca 314
 Val Asp Arg Leu Arg His His Leu Leu Pro Met Tyr Ser Tyr Asp Pro
 15 20 25
 gct gag gaa ctg crt gag gct gag cag gag ctg ctc tct gac atg gga 362
 Ala Glu Glu Leu Xaa Glu Ala Glu Gln Glu Leu Leu Ser Asp Met Gly
 30 35 40
 gac ccc aag gtg gta cat ggc tgg cag agt ggc tac cag cac aag cgg 410
 Asp Pro Lys Val Val His Gly Trp Gln Ser Gly Tyr Gln His Lys Arg
 45 50 55
 atg cca ctg ctg gat gtc aag acg tgacctgacc cccttgcccc acccttcaga 464
 Met Pro Leu Leu Asp Val Lys Thr
 60 65
 gcctggggtc ctgga 479

<210> 362
 <211> 453
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 30..440

<220>
 <221> sig_peptide
 <222> 30..83
 <223> Von Heijne matrix
 score 7
 seq VLGLVLLSVTVQG/KV

[illegible]

```
<210> 363
<211> 417
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 82..384
```

```
<220>
<221> sig_peptide
<222> 82..168
<223> Von Heijne matrix
      score 6.900000009536743
      seq VCMCVFSTQALG/EM
```

```
<220>
<221> misc_feature
<222> 404
<223> n=a, g, c o r t
```

<400> 363																
gctacctcct ggtgtctgtt gcttgggtgtg tgggaaaccc ccacacgctg gaccacagaa																60
gtcttctgtg ttgactgttg a atg agg gaa gag aaa aag ccc ttt gag aga																111
Met Arg Glu Glu Lys Lys Pro Phe Glu Arg																
-25 -20																
gag aga gag agt gtg tgt gtg tgt atg tgt gtg ttt tcc act caa gga																159
Glu Arg Glu Ser Val Cys Val Cys Met Cys Val Phe Ser Thr Gln Gly																
-15 -10 -5																
gct ttg ggg gaa atg gct gca cac ttc ata gat gaa aag ctg agg ccc																207
Ala Leu Gly Glu Met Ala Ala His Phe Ile Asp Glu Lys Leu Arg Pro																

206

```

      1           5           10
agt gag ggg aat ggt cac aga ggg aca ctg gat agc ctg agt tcg gac      255
Ser Glu Gly Asn Gly His Arg Gly Thr Leu Asp Ser Leu Ser Ser Asp
      15           20           25
caa gag tcc tac atc ccc tcc acc gcc gac ccc acc cag gct ggc cct      303
Gln Glu Ser Tyr Ile Pro Ser Thr Ala Asp Pro Thr Gln Ala Gly Pro
      30           35           40           45
gag ctg ctg cac aag aac ttg ccc gtg acc tcc cga tcc cag ccg ctg      351
Glu Leu Leu His Lys Asn Leu Pro Val Thr Ser Arg Ser Gln Pro Leu
      50           55           60
ccc tct gac ctc gcg atc cca gcc gct gcc ctc tgacctcccg atcccagccn      404
Pro Ser Asp Leu Ala Ile Pro Ala Ala Ala Leu
      65           70
gctgccctcg ctg      417

```

<210> 364
 <211> 529
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 114..497

<220>
 <221> sig_peptide
 <222> 114..191
 <223> Von Heijne matrix
 score 6.80000019073486
 seq VFLSLFLIQLLIS/FS

```

<400> 364
ctcagcacccc agggcggtgg taggtcacag tctctgggcg ggtctcagtg tccaacactg      60
tagctggtgc ctgccaggtt cccagtggtt ggggtcacca ggtctgaaga gag atg      116
                                     Met
tgc tgg ctg cgg gca tgg ggc cag atc ctc ctg cca gtt ttc ctc tcc      164
Cys Trp Leu Arg Ala Trp Gly Gln Ile Leu Leu Pro Val Phe Leu Ser
-25           -20           -15           -10
ctc ttt ctc atc caa ttg ctt atc agc ttc tca gag aat ggt ttt atc      212
Leu Phe Leu Ile Gln Leu Leu Ile Ser Phe Ser Glu Asn Gly Phe Ile
      -5           1           5
cac agc ccc agg aac aat cag aaa cca aga gat ggg aat gaa gag gaa      260
His Ser Pro Arg Asn Asn Gln Lys Pro Arg Asp Gly Asn Glu Glu Glu
      10           15           20
tgt gct gta aag aag agt tgt caa ttg tgc aca gaa gat aag aaa tgt      308
Cys Ala Val Lys Lys Ser Cys Gln Leu Cys Thr Glu Asp Lys Lys Cys
      25           30           35
gtt tgg tgt agt gaa gaa aaa gca tgc aaa aaa tac tgt ttt ccc tat      356
Val Trp Cys Ser Glu Glu Lys Ala Cys Lys Lys Tyr Cys Phe Pro Tyr
      40           45           50           55
ttc ggt tgt cga ttc agt tct ata tat tgg tta aac tgt aaa gtt gac      404
Phe Gly Cys Arg Phe Ser Ser Ile Tyr Trp Leu Asn Cys Lys Val Asp
      60           65           70
atg ttt gga atc atg atg ctt cta ctc att gca gta tta att aca gga      452
Met Phe Gly Ile Met Met Leu Leu Ile Ala Val Leu Ile Thr Gly
      75           80           85
ttc gtt tgg tac tgc tgc gcc tat cac ttt tac ctg cag gat ata      497
Phe Val Trp Tyr Cys Cys Ala Tyr His Phe Tyr Leu Gln Asp Ile
      90           95           100
tgatgaatag ataattgraa agagatcctc ca      529

```

<210> 365
 <211> 505

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 71..496

<220>
<221> sig_peptide
<222> 71..142
<223> Von Heijne matrix
score 6.5
seq LSICLSAVATATG/AE

<400> 365
gacgaggacs cgctgcgca gaggcggcag caccaccggg gttgactccg ggggcgcggc 60
gaggagagac atg agg ctg agc tgg ttc cgg gtc ctg aca gta ctg tcc 109
Met Arg Leu Ser Trp Phe Arg Val Leu Thr Val Leu Ser
-20 -15
atc tgc ctg agc gcc gtg gcc acg gcc acg ggg gcc gag ggc aaa agg 157
Ile Cys Leu Ser Ala Val Ala Thr Ala Thr Gly Ala Glu Gly Lys Arg
-10 -5 1 5
aag ctg cag atc ggg gtc aag aag cgg gtg gac cac tgt ccc atc aaa 205
Lys Leu Gln Ile Gly Val Lys Lys Arg Val Asp His Cys Pro Ile Lys
10 15 20
tcg cgc aaa ggg gat gtc ctg cac atg cac tac acg ggg aag ctg gaa 253
Ser Arg Lys Gly Asp Val Leu His Met His Tyr Thr Gly Lys Leu Glu
25 30 35
gat ggg aca gag ttt gac agc agc ctg ccc cag aac cag ccc ttt gtc 301
Asp Gly Thr Glu Phe Asp Ser Ser Leu Pro Gln Asn Gln Pro Phe Val
40 45 50
ttc tcc ctt ggc aca ggc cag gtc atc aag ggc tgg gac cag ggg ctg 349
Phe Ser Leu Gly Thr Gly Gln Val Ile Lys Gly Trp Asp Gln Gly Leu
55 60 65
ctg ggg atg tgt gag ggg gaa aag cgc aas ykg gtg atc cca tcc gag 397
Leu Gly Met Cys Glu Gly Glu Lys Arg Xaa Xaa Val Ile Pro Ser Glu
70 75 80 85
cta ggg tat gga gag cgg gga gct ccc cca aag att cca ggc ggt gca 445
Leu Gly Tyr Gly Glu Arg Gly Ala Pro Pro Lys Ile Pro Gly Gly Ala
90 95 100
acc ctg gtg ttc gag gtg gag ctg ctc aaa ata gag cga cga act gag 493
Thr Leu Val Phe Glu Val Glu Leu Leu Lys Ile Glu Arg Arg Thr Glu
105 110 115
ctg taaccagac 505
Leu

<210> 366
<211> 539
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 65..274

<220>
<221> sig_peptide
<222> 65..136
<223> Von Heijne matrix
score 6.5
seq ILLTRPACLGSWA/EI

<400> 366

208

```

cactgattct ttggccatcg ctactacaat tttagaattg ctggcaattt tttctgcaca      60
ctgg atg gct tct tcc acc agt gtc tca aca gga caa atc ttg ctg aca      109
      Met Ala Ser Ser Thr Ser Val Ser Thr Gly Gln Ile Leu Leu Thr
            -20                      -15                      -10
aga cct gct tgc ttg ggg tcc tgg gct gag atc cgg tca ccg gtg agg      157
Arg Pro Ala Cys Leu Gly Ser Trp Ala Glu Ile Arg Ser Pro Val Arg
            -5                      1                      5
acc atc tcc atc gcc agc gac ttc cca aca gca cgg gtg agt ctc tgg      205
Thr Ile Ser Ile Ala Ser Asp Phe Pro Thr Ala Arg Val Ser Leu Trp
            10                      15                      20
gtg ccg ccc gca cct ggg atg gtt cct att aag atc tcc ggc tgt gca      253
Val Pro Pro Ala Pro Gly Met Val Pro Ile Lys Ile Ser Gly Cys Ala
            25                      30                      35
aac tgg gcc ttc tca ccg gca tagatgatat cacacatcat ggcaagctca      304
Asn Trp Ala Phe Ser Pro Ala
            40                      45
cagccccgcg caaaggcata gccattgaca gcagcgatga ctggattctt gacctgggtg      364
agtggtgcc agtgcttcaa gaacttgctg gagtaacagt cctggaaact caggttctgc      424
atttccttga tatcagctct agtgaaaagg gcggtagttg gtggtggaac cagaaacgga      484
cgccggtgct tkgagcgggt cttaaattat atttaaaaag taactttttg tataa      539

```

<210> 367
 <211> 460
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 14..178

<220>
 <221> sig_peptide
 <222> 14..82
 <223> Von Heijne matrix
 score 6.40000009536743
 seq LGALSGAAALGFA/SY

```

<400> 367
ctgtttctaca gct atg gcc ggg cca gct gca gct ttc cgc cgc ttg ggc      49
      Met Ala Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly
            -20                      -15
gcc ttg tcc gga gct gcg gcc tta ggc ttc gct tcc tac ggg gcg cac      97
Ala Leu Ser Gly Ala Ala Leu Gly Phe Ala Ser Tyr Gly Ala His
            -10                      -5                      1                      5
ggc gcm aat tcc cag atg cct acg gga agg agc tgt ttg aca agg cca      145
Gly Ala Asn Ser Gln Met Pro Thr Gly Arg Ser Cys Leu Thr Arg Pro
            10                      15                      20
aca aac acc act tct tac aca gcc tgg ccc tgt taggggtgcc ccattgcaga      198
Thr Asn Thr Thr Ser Tyr Thr Ala Trp Pro Cys
            25                      30
aagccactct gggctggggt attgctagct tccggaacga ccttattctg caccagcttt      258
tactaccagg ctctgagtgg agaccccagc atccagactt tggcccctgc gggagggacc      318
ctgctactct tgggctggct tgccttggtc ctttgagctc ccttttgctt aattactggg      378
ttttctgggc agtttttttt ttaaagagtt ggagtaagaa gaggattaaa aaggaaaggc      438
aaataaaaaa aaaaaaaaaa gc      460

```

<210> 368
 <211> 482
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS

<222> 14..178

<220>

<221> sig_peptide

<222> 14..82

<223> Von Heijne matrix

score 6.40000009536743

seq LGALSGAAALGFA/SY

<400> 368

```

ctgttctaca gct atg gcc ggg cca gct gca gct ttc cgc cgc ttg ggc      49
              Met Ala Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly
              -20                      -15
gcc ttg tcc gga gct gcg gcc tta ggc ttc gct tcc tac ggg gcg cac      97
Ala Leu Ser Gly Ala Ala Ala Leu Gly Phe Ala Ser Tyr Gly Ala His
-10                      -5                      1                      5
ggc gcm aat tcc cag atg cct acg gga agg agc tgt ttg aca agg cca      145
Gly Ala Asn Ser Gln Met Pro Thr Gly Arg Ser Cys Leu Thr Arg Pro
              10                      15                      20
aca aac acc act tct tac aca gcc tgg ccc tgt taggggtgcc ccattgcaga      198
Thr Asn Thr Thr Ser Tyr Thr Ala Trp Pro Cys
              25                      30
aagccactct gggctggggt attgctagct tccggaacga ccttattctg caccagcttt      258
tactaccagg ctctgagtgg agacccagc atccagactt tggcccctgc gggagggacc      318
ctgctactct tgggctggct tgccttggct ctttgagctc ctttttgctt aattactggg      378
ttttctgggc agtttttttt ttaaagagtt ggagtaagaa gaggattaaa aaggaaaggc      438
aaataaactt tggagtcttt gttcatcaaa aaaaaaaaaa aaaa                        482

```

<210> 369

<211> 497

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 161..322

<220>

<221> sig_peptide

<222> 161..211

<223> Von Heijne matrix

score 6.40000009536743

seq LTFAQLLFATVLG/IA

<400> 369

```

agatgggggt aggagctctc cagacagtgt cccctcactc acccccgcgg cgctggcggt      60
cttctcctgg gtgttgaagt acaagatttg aagagatttc ctgggtgtgg agtgtgactt      120
tccaaaacca gcttttcctt gagctgtatt tgttgcagca atg ttt agg aga ttg      175
              Met Phe Arg Arg Leu
              -15
act ttt gca caa ctg ctt ttt gcc act gtc ctt gga att gct gga gga      223
Thr Phe Ala Gln Leu Leu Phe Ala Thr Val Leu Gly Ile Ala Gly Gly
-10                      -5                      1
gta tat att ttt caa cca gta ttt gaa cag tat gcc aaa gat cag aag      271
Val Tyr Ile Phe Gln Pro Val Phe Glu Gln Tyr Ala Lys Asp Gln Lys
5                      10                      15                      20
gaa tta aaa gaa aag atg cag ttg gta caa gaa tca gaa gag aag aaa      319
Glu Leu Lys Glu Lys Met Gln Leu Val Gln Glu Ser Glu Glu Lys Lys
              25                      30                      35
agt taatactaca tggagttagg cctggcgagc tggctcacgc ctgtaatccc      372
Ser
agcacttttg gaggccgagg cgggtggatc aggtggtcag gagttcaaga ccagcctgac      432
caacatggtg aaacctgtct ctactgaaaa taaaaaatc agccaggctt ggtggcatgc      492

```

gctgt

497

<210> 370
 <211> 422
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 47..295

<220>
 <221> sig_peptide
 <222> 47..91
 <223> Von Heijne matrix
 score 6.30000019073486
 seq LFLSLPVLVVVLS/IV

<400> 370
 aaccaagccc tccagcaagg attcagagtg cccctccggc ctcgcc atg agg ctc 55
 Met Arg Leu
 -15
 ttc ctg tcg ctt ccg gtc ctg gtg gtg gtt ctg tcg atc gtc ttg gaa 103
 Phe Leu Ser Leu Pro Val Leu Val Val Val Leu Ser Ile Val Leu Glu
 -10 -5 1
 ggc cca gcc cca gcc cag ggg acc cca gac gtc tcc agt gcc ttg gat 151
 Gly Pro Ala Pro Ala Gln Gly Thr Pro Asp Val Ser Ser Ala Leu Asp
 5 10 15 20
 aag ctg aag gag ttt gga aac aca ctg gag gac aag gct cgg gaa ctc 199
 Lys Leu Lys Glu Phe Gly Asn Thr Leu Glu Asp Lys Ala Arg Glu Leu
 25 30 35
 atc agc cgc atc aaa cag agt gaa ctt tct gcc aag atg cgg gag tgg 247
 Ile Ser Arg Ile Lys Gln Ser Glu Leu Ser Ala Lys Met Arg Glu Trp
 40 45 50
 ttt tca gag aca ttt cag aaa gtg aag gat aaa ctc aag att gac tca 295
 Phe Ser Glu Thr Phe Gln Lys Val Lys Asp Lys Leu Lys Ile Asp Ser
 55 60 65
 tgaggacctg aagggtgaca tcccaggagg ggcctctgaa atttcccaca cccagcgcc 355
 tgtgctgagg actccctcca tgtggcccca ggtgccacca ataaaaatcc tacagaaaac 415
 taatata 422

<210> 371
 <211> 495
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 66..359

<220>
 <221> sig_peptide
 <222> 66..194
 <223> Von Heijne matrix
 score 6
 seq FLIYVALLRVSEC/LP

<400> 371
 gggcgcgagg cggccaccgt ggagagcaga gcgcggcggc tggaagctgc taagtcagag 60
 ccgcg atg ttc cgg att gag ggc ctc gcg ccg aag ctg gac ccg gag gag 110
 Met Phe Arg Ile Glu Gly Leu Ala Pro Lys Leu Asp Pro Glu Glu
 -40 -35 -30
 atg aaa cgg aag atg cgc gag gat gtg atc tcc tcc ata cgg aac ttt 158

211

```

Met Lys Arg Lys Met Arg Glu Asp Val Ile Ser Ser Ile Arg Asn Phe
      -25      -20      -15
ctc atc tac gtg gcc ctc ctg cga gtc agt gag tgt ctc ccg ggc tgt      206
Leu Ile Tyr Val Ala Leu Leu Arg Val Ser Glu Cys Leu Pro Gly Cys
      -10      -5      1
gac tgt gat acc agc ggg gag ctc acc gac ggg cac ccc tta act cta      254
Asp Cys Asp Thr Ser Gly Glu Leu Thr Asp Gly His Pro Leu Thr Leu
5      10      15      20
agg ggt cat cgg ggc ctt cga act gag ctg aac ggt agc ggg gag caa      302
Arg Gly His Arg Gly Leu Arg Thr Glu Leu Asn Gly Ser Gly Glu Gln
      25      30      35
gga ggc tcc att tat ctt aaa gaa att gga cag cat atg aag aca gga      350
Gly Gly Ser Ile Tyr Leu Lys Glu Ile Gly Gln His Met Lys Thr Gly
      40      45      50
cat cac ata tgaatgcacg atatgaagag cctgggttaca gtttcgactc      399
His His Ile
      55
ctctctgcaa gtgaataggc ccagaaaggt gtaagagact ctttgaatgg acataaaatt      459
ctgcttggtta agaacaagtt tggctctggt aactga      495

```

<210> 372

<211> 515

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 115..360

<220>

<221> sig_peptide

<222> 115..243

<223> Von Heijne matrix

score 5.59999990463257

seq VVSFALIATLVYA/LF

<400> 372

```

ataaggggac gtctagtggg ttgcccggga ggggtggcggg aggggtcctgg aaataatctg      60
tcctctgtcg ccgggaactg gcgaggtagt tccttcgcgg tggagagacc tgga atg      117
                                     Met
gcc aaa tat caa ggt gaa gtt caa agt ttg aaa ctg gat gat gat tca      165
Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp Ser
      -40      -35      -30
gtt ata gaa gga gta agc gac caa gta ctt gtg gca gtt gtg gtc agt      213
Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val Ser
      -25      -20      -15
ttc gct ttg att gct acc ctg gta tat gca ctt ttc aga aat gta cat      261
Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val His
      -10      -5      1      5
caa aac att cac cca gaa aac cag gag cta gta agg gta ctt cga gaa      309
Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg Glu
      10      15      20
cag ctt caa aca gaa cag gat gca cct gct gac tcg aca gca gtt cta      357
Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val Leu
      25      30      35
cac tgacatgtac tgtcccatct gctgcacca agcctccttc ccggtggaga      410
His
ccaactgtgg acatcttttt tgtgggtgct gcattattgc ttactggcga tatgggttcat      470
ggcttggggc aatcagttgt ccaatctgta gacaaacgag acatg      515

```

<210> 373

<211> 513

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 56..352

<220>

<221> sig_peptide

<222> 56..139

<223> Von Heijne matrix

score 5.5

seq LGYLVLSEGAFLA/SS

<400> 373

```

ctgaagccg aagctaccta tctggtaggg agctcccca gcaccgaaga ctgcg atg      58
                                     Met
act tct gca ctg acc cag ggg ctg gag cga atc cca gac cag ctc ggc      106
Thr Ser Ala Leu Thr Gln Gly Leu Glu Arg Ile Pro Asp Gln Leu Gly
      -25                -20                -15
tac ctg gta ctg agt gaa ggt gca gtg ctg gcg tca tct ggg gac ctg      154
Tyr Leu Val Leu Ser Glu Gly Ala Val Leu Ala Ser Ser Gly Asp Leu
      -10                -5                1                5
gag aat gat gag cag gca gcc agt gcc atc tct gag ctg gtc agc aca      202
Glu Asn Asp Glu Gln Ala Ala Ser Ala Ile Ser Glu Leu Val Ser Thr
                        10                15                20
gcc tgc ggt ttc cgg ctg cac cgc ggc atg aat gtg ccc ttc aag cgc      250
Ala Cys Gly Phe Arg Leu His Arg Gly Met Asn Val Pro Phe Lys Arg
                        25                30                35
ctg tct gtg gtc ttt gga gaa cac aca ctg ctg gtg acg gtg tca gga      298
Leu Ser Val Val Phe Gly Glu His Thr Leu Leu Val Thr Val Ser Gly
                        40                45                50
cag agg gtg ttt gtg gtg aag agg cag aac cga ggt cgg gag ccc att      346
Gln Arg Val Phe Val Val Lys Arg Gln Asn Arg Gly Arg Glu Pro Ile
                        55                60                65
gat gtc tgagcctgcc ggaggcgagg gtcggagaag cggattgggt cctgggcctc      402
Asp Val
70
tgtgatgagg caggcacacc tgtcggtctt ggcttgctgc tagaaactag ggccttctgc      462
tcgccacact cccacccctt acctggacgg gccaggctt ggggaactctg a      513

```

<210> 374

<211> 632

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 2..559

<220>

<221> sig_peptide

<222> 2..46

<223> Von Heijne matrix

score 5.30000019073486

seq VMAASMARGGVSA/RV

<400> 374

```

g atg cat gtc atg gcc gcc tcc atg gcc cgg gga ggc gtg agt gcc agg      49
Met His Val Met Ala Ala Ser Met Ala Arg Gly Gly Val Ser Ala Arg
      -15                -10                -5                1
gtt cta ctg cag gct gcc agg ggc acc tgg tgg aac aga cct ggg ggc      97
Val Leu Leu Gln Ala Ala Arg Gly Thr Trp Trp Asn Arg Pro Gly Gly
                        5                10                15

```

213

```

act tcc ggg tcg ggg gag ggg gtg gcg ctg ggg aca acc aga aag ttt      145
Thr Ser Gly Ser Gly Glu Gly Val Ala Leu Gly Thr Thr Arg Lys Phe
      20                      25                      30
caa gcg aca ggc tcg cgc ccg gct gga gag gag gac gcg ggc ggc ccg      193
Gln Ala Thr Gly Ser Arg Pro Ala Gly Glu Glu Asp Ala Gly Gly Pro
      35                      40                      45
gag cgg ccc ggg gac gtg gtg aac gtg gtg ttc gta gac cgc tca ggc      241
Glu Arg Pro Gly Asp Val Val Asn Val Val Phe Val Asp Arg Ser Gly
      50                      55                      60
cag cgg atc cca gtg agt ggc aga gtc ggg gac aat gtt ctt cac ctg      289
Gln Arg Ile Pro Val Ser Gly Arg Val Gly Asp Asn Val Leu His Leu
      70                      75                      80
gcc cag cgc cac ggg gtg gac ctg gaa ggg gcc tgt gaa gcc tcc ctg      337
Ala Gln Arg His Gly Val Asp Leu Glu Gly Ala Cys Glu Ala Ser Leu
      85                      90                      95
gcc tgc tcc acc tgc cat gtg tat gtg agt gaa gac cac ctg gat ctc      385
Ala Cys Ser Thr Cys His Val Tyr Val Ser Glu Asp His Leu Asp Leu
      100                      105                      110
ctg cct cct ccc gag gag agg gaa gac gac atg cta gac atg gcc ccc      433
Leu Pro Pro Pro Glu Glu Arg Glu Asp Asp Met Leu Asp Met Ala Pro
      115                      120                      125
ctc ctc cag gag aac tcg cgg ctg ggc tgc cag att gtg ctg aca ccg      481
Leu Leu Gln Glu Asn Ser Arg Leu Gly Cys Gln Ile Val Leu Thr Pro
      130                      135                      140                      145
gag ctg saa gga gcg gaa ttc acc ctg ccc aag atc acc agg aac ttc      529
Glu Leu Xaa Gly Ala Glu Phe Thr Leu Pro Lys Ile Thr Arg Asn Phe
      150                      155                      160
tac gtg gat ggc cat gtc ccc aag ccc cac tgacatgaac acctggacca      579
Tyr Val Asp Gly His Val Pro Lys Pro His
      165                      170
ttccacattg ccatggcccc agggccagat tgagggaata gcaggtgcag cct      632

```

<210> 375

<211> 503

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 59..427

<220>

<221> sig_peptide

<222> 59..202

```

<223> Von Heijne matrix
      score 5.19999980926514
      seq LLMLLLFLSELQY/YL

```

<400> 375

```

aggccggaag agggagtctg taggggcggg ccgggcgtcc cctttccggc cggtcccc      58
atg gag gcg ctg ggg aag ctg aag cag ttc gat gcc tac ccc aag act      106
Met Glu Ala Leu Gly Lys Leu Lys Gln Phe Asp Ala Tyr Pro Lys Thr
      -45                      -40                      -35
ttg gag gac ttc cgg gtc aag acc tgc ggg ggc gcc acc gtg acc att      154
Leu Glu Asp Phe Arg Val Lys Thr Cys Gly Gly Ala Thr Val Thr Ile
      -30                      -25                      -20
gtc agt ggc ctt ctc atg ctg cta ctg ttc ctg tcc gag ctg cag tat      202
Val Ser Gly Leu Leu Met Leu Leu Leu Phe Leu Ser Glu Leu Gln Tyr
      -15                      -10                      -5
tac ctc acc acg gag gtg cat cct gag ctc tac gtg gac aag tcg cgg      250
Tyr Leu Thr Thr Glu Val His Pro Glu Leu Tyr Val Asp Lys Ser Arg
      1                      5                      10                      15
gga gat aaa ctg aag atc aac atc gat gta ctk ttt ccg cac atg cct      298

```

214

```

Gly Asp Lys Leu Lys Ile Asn Ile Asp Val Leu Phe Pro His Met Pro
      20      25      30
tgt gcc tat ctg agt att gat gcc atg gat gtg gcc gga gaa cag cag      346
Cys Ala Tyr Leu Ser Ile Asp Ala Met Asp Val Ala Gly Glu Gln Gln
      35      40      45
ctg gat gtg gaa cac aac ctg ttc aag caa cga cta gat aaa gat ggc      394
Leu Asp Val Glu His Asn Leu Phe Lys Gln Arg Leu Asp Lys Asp Gly
      50      55      60
atc ccc gtg agc tca gag gct gag cgg cat gat taatgcagga caccagtgga      447
Ile Pro Val Ser Ser Glu Ala Glu Arg His Asp
      65      70      75
accaacataa tttgctatag gactggatgt ggagtggagg agagcaaagt catggt      503

```

<210> 376

<211> 521

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 145..375

<220>

<221> sig_peptide

<222> 145..231

```

<223> Von Heijne matrix
      score 5.09999990463257
      seq LLLKLQRLPQAEP/VE

```

<400> 376

```

gacaaacttg ttctgcgggc tgcgcatggg tgcgaggggtg gaatctcggt gctgcgacga      60
gtgtggggcc arccgtggag gctccaggtg ttctctctgc cccagcagag cccggcagga      120
gccccaacag gaagccagcg cggc atg gct gcc acc gac ttc gtg cag gag      171
                        Met Ala Ala Thr Asp Phe Val Gln Glu
                        -25
atg cgc gcc gtg ggc gag agg ctg ctg ctc aag ctg cag aga ctg ccc      219
Met Arg Ala Val Gly Glu Arg Leu Leu Leu Lys Leu Gln Arg Leu Pro
-20      -15      -10      -5
cag gct gag ccc gtg gag atc gtg gcc ttc tca gtc atc atc ctt ttc      267
Gln Ala Glu Pro Val Glu Ile Val Ala Phe Ser Val Ile Ile Leu Phe
      1      5      10
aca gct act gtt ctg ctg ttg ctg ctg ata gcc tgc agc tgc tgc tgc      315
Thr Ala Thr Val Leu Leu Leu Leu Leu Ile Ala Cys Ser Cys Cys Cys
      15      20      25
act cac tgc tgc tgc cct gag crg aga ggc agg aag gtc cag gtg cag      363
Thr His Cys Cys Cys Pro Glu Xaa Arg Gly Arg Lys Val Gln Val Gln
      30      35      40
ccg aca cca cca tgacggackg gcgatggctg aggagaagct ggagaggaga      415
Pro Thr Pro Pro
45
tgGCCaatgc catgacacag gccatcagcc tggccctgca gcccttacct ctcaagacsa      475
ggctcccttg cccagctctg gccagccag gtacctggac actgac      521

```

<210> 377

<211> 461

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 17..337

<220>

<221> sig_peptide

<222> 17..97

<223> Von Heijne matrix

score 4.90000009536743

seq VLSTLSLVQFSPS/GR

<400> 377

```

acttttagctg ctgcag atg ctg ttt gaa gaa gct ctg ccc ctc agc tgc tct      52
                Met Leu Phe Glu Glu Ala Leu Pro Leu Ser Cys Ser
                -25                               -20
gat cct gtg ctt agc act ctt agc ctg gtg cag ttc agc ccc agt gga      100
Asp Pro Val Leu Ser Thr Leu Ser Leu Val Gln Phe Ser Pro Ser Gly
-15                               -10                -5                1
agg acc cag gac ctg ctc tct cca ggg gtg gag aac ctg tcg gtg ctg      148
Arg Thr Gln Asp Leu Leu Ser Pro Gly Val Glu Asn Leu Ser Val Leu
                5                               10                15
gac gtg tcc cct ctg ggc ttg gcc tgc tgt ctg ctc act ctc acc atg      196
Asp Val Ser Pro Leu Gly Leu Ala Cys Cys Leu Leu Thr Leu Thr Met
                20                               25                30
tcc tgc cca ggg cct gac cct cct gag ggg ccc ggg acc cag cgt gtg      244
Ser Cys Pro Gly Pro Asp Pro Pro Glu Gly Pro Gly Thr Gln Arg Val
                35                               40                45
tgg caa ggg gct cta cgg atc cta cag ctc cca gga gcc cca gat ggg      292
Trp Gln Gly Ala Leu Arg Ile Leu Gln Leu Pro Gly Ala Pro Asp Gly
50                               55                60                65
gtt tca cca tat cag cca gtc tgg tct cga act cct gac ctc aag      337
Val Ser Pro Tyr Gln Pro Val Trp Ser Arg Thr Pro Asp Leu Lys
                70                               75                80
tgattcgcca gcctcgctcct cccaaagtgc tgggattaca gatgtgagcc accatgcccc      397
tggagaatgt attacttttg taattttaaaa aataatagta ataaaataaa gatctttaca      457
atgt                                                                461

```

<210> 378

<211> 495

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 148..483

<220>

<221> sig_peptide

<222> 148..264

<223> Von Heijne matrix

score 4.80000019073486

seq IILFSAIVGFIYG/YV

<220>

<221> misc_feature

<222> 43,112,122,154

<223> n=a, g, c or t

<400> 378

```

cctaatacgaa aattcgctct cccgggctta gaaggcccggt ctnctgacgc gcagtgccag      60
accttacccc tcacggctcct taagtctcgg tcgccctcgc ctgcgagcct gncacccgcg      120
mncagctgcc cgcctctctca gccagcc atg ctg nag cat ctg agc tcg ctg ccc      174
                Met Leu Xaa His Leu Ser Ser Leu Pro
                -35
acg cag atg gat tac aag ggc cag aag cta gct gaa cag atg ttt cag      222
Thr Gln Met Asp Tyr Lys Gly Gln Lys Leu Ala Glu Gln Met Phe Gln
-30                -25                -20                -15
gga att att ctt ttt tct gca ata gtt gga ttt atc tac ggg tac gtg      270

```


216

Gly	Ile	Ile	Leu	Phe	Ser	Ala	Ile	Val	Gly	Phe	Ile	Tyr	Gly	Tyr	Val	
			-10						-5					1		
gct	gaa	cag	ttc	ggg	tgg	act	gtc	tat	ata	gtt	atg	gcc	gga	ttt	gct	318
Ala	Glu	Gln	Phe	Gly	Trp	Thr	Val	Tyr	Ile	Val	Met	Ala	Gly	Phe	Ala	
		5					10				15					
ttt	tca	tgt	ttg	ctg	aca	ctt	cct	cca	tgg	ccc	atc	tat	cgc	cgg	cat	366
Phe	Ser	Cys	Leu	Leu	Thr	Leu	Pro	Pro	Trp	Pro	Ile	Tyr	Arg	Arg	His	
	20				25					30						
cct	ctc	aag	tgg	tta	cct	gtt	caa	gca	cag	acg	aca	aga	aac	cag	ggg	414
Pro	Leu	Lys	Trp	Leu	Pro	Val	Gln	Ala	Gln	Thr	Thr	Arg	Asn	Gln	Gly	
	35				40				45					50		
aaa	gaa	aaa	tta	aga	ggc	atg	cta	aaa	ata	att	gag	gtt	ttc	atg	att	462
Lys	Glu	Lys	Leu	Arg	Gly	Met	Leu	Lys	Ile	Ile	Glu	Val	Phe	Met	Ile	
			55				60						65			
cag	cac	ctg	ctt	ttg	ttt	ctg	tgagatgagc	ta								495
Gln	His	Leu	Leu	Leu	Phe	Leu										
			70													

<210> 379

<211> 541

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 60..368

<220>

<221> sig_peptide

<222> 60..122

<223> Von Heijne matrix

score 4.69999980926514

seq ALRSLNLAPPTVA/AP

<400> 379

acagtcaacg	tcattctagga	gcaccgagca	gcttggctaa	aagtaagggt	gtcgtgctg		59
atg gcc ctg tgc gca ctg acc cgc gct ctg cgc tct ctg aac ctg gcg						107	
Met Ala Leu Cys Ala Leu Thr Arg Ala Leu Arg Ser Leu Asn Leu Ala							
-20		-15		-10			
ccc ccg acc gtc gcc gcc cct gcc ccg agt ctg ttc ccc gcc gcc cag					155		
Pro Pro Thr Val Ala Ala Pro Ala Pro Ser Leu Phe Pro Ala Ala Gln							
-5		1		5	10		
atg atg aac aat ggc ctc ctc caa cag ccc tct gcc ttg atg ttg ctc					203		
Met Met Asn Asn Gly Leu Leu Gln Gln Pro Ser Ala Leu Met Leu Leu							
	15		20		25		
ccc tgc cgc cca gtt ctt act tct gtg gcc ctt aat gcc aac ttt gtg					251		
Pro Cys Arg Pro Val Leu Thr Ser Val Ala Leu Asn Ala Asn Phe Val							
	30		35		40		
tcc tgg aag agt cgt acc aag tac acc att aca cca gtg aag atg agg					299		
Ser Trp Lys Ser Arg Thr Lys Tyr Thr Ile Thr Pro Val Lys Met Arg							
	45		50		55		
aag tct ggg ggc cga gac cac aca ggt gct gga aac gtg cgt agc aac					347		
Lys Ser Gly Gly Arg Asp His Thr Gly Ala Gly Asn Val Arg Ser Asn							
	60		65		70	75	
agt agg ccg agt atc caa cgt tgatcataac aaacgggtca ttggcaaggc					398		
Ser Arg Pro Ser Ile Gln Arg							
	80						
aggtcgcaac cgctggctgg gcaagaggcc taacagtggg cggtggcacc gcaagggggg					458		
ctgggctggc cgaaagattc ggccactacc ccccatgaag agttacgtga agctgccttc					518		
tgcttctgcc aaagctgata tcc					541		

<210> 380

<211> 545

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 80..355

<220>

<221> sig_peptide

<222> 80..124

<223> Von Heijne matrix

score 4.59999990463257

seq VLAGSLLGPTSRS/AA

<400> 380

```

cttttccttt gttccagggc tggagcggct ctgggctccg gaatcgcccg cagccggtac      60
tgcgggaccc actgcggat atg gct gtc ttg gct gga tcc ctg ttg ggc ccc      112
                Met Ala Val Leu Ala Gly Ser Leu Leu Gly Pro
                -15                -10                -5
acg agt agg tcg gca gcg ttg ctg ggt ggc agg tgg ctc cag ccc cgg      160
Thr Ser Arg Ser Ala Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro Arg
                1                5                10
gcc tgg ctg ggg ttc cca gac gcc tgg ggc ctc ccc acc ccg cag cag      208
Ala Trp Leu Gly Phe Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln
                15                20                25
gcc cgg ggc aag gct cgc ggg aat gag tat cag ccg agc aac atc aaa      256
Ala Arg Gly Lys Ala Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys
                30                35                40
cgc aag aac aag cac ggc tgg gtc cgg cgc ctg agc acg ccg gcc ggc      304
Arg Lys Asn Lys His Gly Trp Val Arg Arg Leu Ser Thr Pro Ala Gly
                45                50                55                60
gtg cag gtc atc ctt cgc cga atg ctc aag ggc cgc aag tcg ctg agc      352
Val Gln Val Ile Leu Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser
                65                70                75
cat tgaggatcgc gacgcagtcg gcgggaccct catggaagca tcgccctcgc      405
His
ctcggacctt gcctggcgct atttttgcag ggagctgggg agcaggaacg ctcggacct      465
gagtgtcttc catattgtgg gggtgaagtc tggatgggag cttgccaagt ccctttttag      525
gctttttaat taggaagcat      545

```

<210> 381

<211> 498

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 137..397

<220>

<221> sig_peptide

<222> 137..253

<223> Von Heijne matrix

score 4.59999990463257

seq LQGARVLLTGANA/GV

<400> 381

```

ggggtgcggt cggggacccg gcaggaggcg gccgagaaga gaggaccgtg ggggcgttcg      60
cgtggctccc agcccgggac cccaccccg ctggacagtg ggggaaactg aggcctgagc      120
gggcccacac aggacc atg aag gtg ctt ctc ctc aca ggg ctg ggg gcc ctg      172
                Met Lys Val Leu Leu Leu Thr Gly Leu Gly Ala Leu
                -35                -30
ttc ttc gcc tat tat tgg gat gac aac ttc gac cca gcc agc ctc cag      220

```

218

```

Phe Phe Ala Tyr Tyr Trp Asp Asp Asn Phe Asp Pro Ala Ser Leu Gln
    -25                -20                -15
gga gcg cga gtg ctg ctg aca ggg gcc aac gct ggt gtt ggt gag gag      268
Gly Ala Arg Val Leu Leu Thr Gly Ala Asn Ala Gly Val Gly Glu Glu
    -10                -5                1                5
ctg gcc tat cac tac gcg cgt ctg ggc tcc cac ctg gtg ctc act gcc      316
Leu Ala Tyr His Tyr Ala Arg Leu Gly Ser His Leu Val Leu Thr Ala
                10                15                20
cac act gag gct ctc ctg cag aag gca cgg tgg ctc acg ctt gta gtc      364
His Thr Glu Ala Leu Leu Gln Lys Ala Arg Trp Leu Thr Leu Val Val
                25                30                35
agc act ttg gga ggc cga gga aag tgg atc acc tgaggtcagg agttcaagac      417
Ser Thr Leu Gly Gly Arg Gly Lys Trp Ile Thr
                40                45
cagcctggcc aacatggcga aacctggtct ctactaagaa gacaaaaatt aaggccaggc      477
atggtggctc atgcctgtaa t      498

```

<210> 382

<211> 459

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 43..417

<220>

<221> sig_peptide

<222> 43..87

<223> Von Heijne matrix

score 4.59999990463257

seq LLTHNLLSSHVRG/VG

<400> 382

```

acctttttcc ggttccggcc tggcgagagt ttgtgcggcg ac atg aaa ctg ctt      54
                                Met Lys Leu Leu
                                -15
acc cac aat ctg ctg agc tcg cat gtg cgg ggg gtg ggg tcc cgt ggc      102
Thr His Asn Leu Leu Ser Ser His Val Arg Gly Val Gly Ser Arg Gly
    -10                -5                1                5
ttc ccc ctg cgc ctc cag gcc acc gag gtc cgt atc tgc cct gtg gaa      150
Phe Pro Leu Arg Leu Gln Ala Thr Glu Val Arg Ile Cys Pro Val Glu
                10                15                20
ttc aac ccc aac ttc gtg gcg cgt atg ata cct aaa gtg gag tgg tcg      198
Phe Asn Pro Asn Phe Val Ala Arg Met Ile Pro Lys Val Glu Trp Ser
                25                30                35
gcg ttc ctg gag gcg gcc gat aac ttg cgt ctg atc cag gtg ccg aaa      246
Ala Phe Leu Glu Ala Ala Asp Asn Leu Arg Leu Ile Gln Val Pro Lys
                40                45                50
ggg ccg gtt gag gga tat gag gag aat gag gag ttt ctg agg acc atg      294
Gly Pro Val Glu Gly Tyr Glu Glu Asn Glu Glu Phe Leu Arg Thr Met
                55                60                65
cac cac ctg ctg ctg gag gtg gaa gtg ata gag ggc acc ctg cag tgc      342
His His Leu Leu Leu Glu Val Glu Val Ile Glu Gly Thr Leu Gln Cys
                70                75                80                85
ccg gaa tct gga cgt atg ttc ccc atc agc cgc ggg atc ccm amc atg      390
Pro Glu Ser Gly Arg Met Phe Pro Ile Ser Arg Gly Ile Pro Xaa Met
                90                95                100
ctg ctg agt gaa gag gaa act gag agt tgattgtgsc aggcgccagt      437
Leu Leu Ser Glu Glu Glu Thr Glu Ser
                105                110
ttttcttgty atgactgtgt at      459

```

<210> 383
 <211> 533
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 82..237

<220>
 <221> sig_peptide
 <222> 82..147
 <223> Von Heijne matrix
 score 4.5
 seq FLLITIALGTKT/ES

<400> 383
 aawttttctca acaatttcctc accgcagagc ctctgaagct cccaccaggc cagctctcct 60
 cccacaacag cttccacag c atg aag atc tcc gtg gct gcc att ccc ttc 111
 Met Lys Ile Ser Val Ala Ala Ile Pro Phe
 -20 -15
 ttc ctc ctc atc acc atc gcc cta ggg acc aag act gaa tcc tcc tca 159
 Phe Leu Leu Ile Thr Ile Ala Leu Gly Thr Lys Thr Glu Ser Ser Ser
 -10 -5 1
 cgg gga cct tac cac ccc tca gag tgc tgc ttc acc tac act acc tac 207
 Arg Gly Pro Tyr His Pro Ser Glu Cys Cys Phe Thr Tyr Thr Thr Tyr
 5 10 15 20
 aag atc ccg cgt cag awt tat gga tta cta tgagaccaac agccagtgtc 257
 Lys Ile Pro Arg Gln Xaa Tyr Gly Leu Leu
 25 30
 ccaagcccg aattgtcttc atcaccaaaa ggggccattc cgtctgtacc aacccagtg 317
 acaagtgggt ccaggactat atcaaggaca tgaaggagaa ctgagtgtacc cagaaggggt 377
 ggcgaaggca cagctcagag acataaagag aagatgccaa ggccccctcc tccaccacc 437
 gctaactctc agccccagtc accctcttgg agcttcctg ctttgaatta aagaccactc 497
 atgctctcca aaaaaaaaaa aaaaaggtca aaagct 533

<210> 384
 <211> 472
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 92..307

<220>
 <221> sig_peptide
 <222> 92..217
 <223> Von Heijne matrix
 score 4.40000009536743
 seq LRLEVAAPGTPA/QP

<400> 384
 aaaagaaatt cccggggcccg gccagcctgt gccgctgtta tgaggagaag caaagcccct 60
 tgcagaagca cttgcggtct gcagcctacg c atg aat agg ttg gca ggt gtg 112
 Met Asn Arg Leu Ala Gly Val
 -40
 ggc tgg cgg gtg gac tac acc ctg agc tcc agc ctg ctg caa tcc gtg 160
 Gly Trp Arg Val Asp Tyr Thr Leu Ser Ser Ser Leu Leu Gln Ser Val
 -35 -30 -25 -20
 gaa gag ccc atg gtg cac ctg cgg ctg gag gtg gca gct gcc cca ggg 208
 Glu Glu Pro Met Val His Leu Arg Leu Glu Val Ala Ala Ala Pro Gly
 -15 -10 -5

220

```

acc cca gcc cag cct gtt gcc atg tcc ctc tca gca gac aag ttc cag      256
Thr Pro Ala Gln Pro Val Ala Met Ser Leu Ser Ala Asp Lys Phe Gln
      1              5              10
gtc ctc ctg gca gaa ctg aag cag gcc cag acc ctg atg agc tcc ctg      304
Val Leu Leu Ala Glu Leu Lys Gln Ala Gln Thr Leu Met Ser Ser Leu
      15              20              25
ggc tgaggagaag ggtgttccag gcctgtgtgg agccgcctg cccgtatgga      357
Gly
30
gtcacgccct ctgaactgct cttcgggagg cagccctggg tctaggatgc tgaggccctg      417
gcccggaactc tggcctccca gatccccagc tgcctcactt ctctcttgag aactt      472

```

<210> 385
 <211> 514
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 113..364

<220>
 <221> sig_peptide
 <222> 113..172
 <223> Von Heijne matrix
 score 4.30000019073486
 seq SLLLSLPPHQGLT/FS

```

<400> 385
ttttttacat ggtgttccca cagctgggag gacacccaca tggtcggcgt gcaggatatt      60
tcgctggacc ctagaaaagc caccacgacc tgtggggccat gatgctaccc ca atg gct      118
                                   Met Ala
                                   -20
gct gct gct gtt cct tct ctt ctt ctt tct ctt cct cct cac cag ggg      166
Ala Ala Ala Val Pro Ser Leu Leu Leu Ser Leu Pro Pro His Gln Gly
      -15              -10              -5
ctc act ttc tcc aac aaa ata caa cct ttt gga gct caa gga gtc ttg      214
Leu Thr Phe Ser Asn Lys Ile Gln Pro Phe Gly Ala Gln Gly Val Leu
      1              5              10
cat ccg gaa cca gga ctg cga gac tgg ctg ctg cca acg tgc tcc aga      262
His Pro Glu Pro Gly Leu Arg Asp Trp Leu Leu Pro Thr Cys Ser Arg
      15              20              25              30
caa ttg cga gtc gca ctg ccg gag aag ggg tcc gag ggc agt ctg tgt      310
Gln Leu Arg Val Ala Leu Pro Glu Lys Gly Ser Glu Gly Ser Leu Cys
      35              40              45
caa acg cag ctg cca gct act cca tgc ttc ctg cct tcg aat acg gtg      358
Gln Thr Gln Leu Pro Ala Thr Pro Cys Phe Leu Pro Ser Asn Thr Val
      50              55              60
aga acg tgaagtcatg agctgctgct aaggcatgtg gcaaccttga agagaaggtc      414
Arg Thr
aagagctacc agccacccaaa agaatgccag cacttcctgt gtctttgctt tggattcatg      474
agaaatatac gttcctatatt gttcaaaaa aaaaaaaaaaw      514

```

<210> 386
 <211> 508
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 83..244

<220>

222

```

tgcctgatct tgcctagaac cggctgcacc tgctgttctc ttgtttatgc aaactggctg 407
cacctgctat tcctttgctt atgcccctac ccctggctat cctaattccc tgttctcctg 467
cctcactatt actgtattct ctacttctaa ataaaaat 505

```

<210> 388
 <211> 775
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 395..736

<220>
 <221> sig_peptide
 <222> 395..490
 <223> Von Heijne matrix
 score 4.19999980926514
 seq LLGSGLFVFSLTA/FN

```

<400> 388
atgatgggta acaggaccgg tgggstcccc aggaagtcct agaggggggc ggggtttggg 60
tggacaagct ttctctgtcc tctcccgaca gagctgacgt gtcctgggtt ccaccgggag 120
cgggcatttc caccggacgg gaggggttcgg ggtgtccggg gctggggaat acgtaggggt 180
tgccgcgcgg tgtggggagt tggggcgtgt ggctgcagtc ccgggagttc ttggaggggg 240
tcggccccacc gagcttccgg accggctgat ctgcccgtag cttgccggag gagggcggas 300
tgactctccg tcccttctcc catcccctcc agtgggtggg acgggcacct cgctggcgct 360
ctcctccctc ctgtccctgc tgctctttgc tggg atg cag atg tac agc cgt cag 415
                                Met Gln Met Tyr Ser Arg Gln
                                -30
ctg gcc tcc acc gag tgg ctc acc atc cag ggc ggc ctg ctt ggt tcg 463
Leu Ala Ser Thr Glu Trp Leu Thr Ile Gln Gly Gly Leu Leu Gly Ser
-25                                -20                                -15                                -10
ggt ctc ttc gtg ttc tgc ctc act gcc ttc aat aat ctg gag aat ctt 511
Gly Leu Phe Val Phe Ser Leu Thr Ala Phe Asn Asn Leu Glu Asn Leu
                                -5                                1                                5
gtc ttt ggc aaa gga ttc caa gca aag atc ttc cct gag att ctc ctg 559
Val Phe Gly Lys Gly Phe Gln Ala Lys Ile Phe Pro Glu Ile Leu Leu
                                10                                15                                20
tgc ctc ctg ttg gct ctc ttt gca tct ggc ctc atc cac cga gtc tgt 607
Cys Leu Leu Leu Ala Leu Phe Ala Ser Gly Leu Ile His Arg Val Cys
                                25                                30                                35
gtc acc acc tgc ttc atc ttc tcc atg gtt ggt ctg tac tac atc aac 655
Val Thr Thr Cys Phe Ile Phe Ser Met Val Gly Leu Tyr Tyr Ile Asn
                                40                                45                                50                                55
aag atc tcc tcc acc ctg tac cag gca gca gct cca gtc ctc aca cca 703
Lys Ile Ser Ser Thr Leu Tyr Gln Ala Ala Ala Pro Val Leu Thr Pro
                                60                                65                                70
gcc aag gtc aca ggc aag agc aag aag aga aac tgaccctgaa tgttcaataa 756
Ala Lys Val Thr Gly Lys Ser Lys Lys Arg Asn
                                75                                80
agttgattct ttgtaaaaaa 775

```

<210> 389
 <211> 466
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 166..324

<220>

```
<221> sig_peptide
```

<222> 166..204

<223> Von Heijne matrix

```
score 4.099999990463257
```

seq MAVAFVLSLGVA A/LY

<400> 389

```

attctgcgcc  tgcgcgcggc  tacagcacgg  ttcgtttttc  ctttagtcag  gaaggacgtt      60
ggtgttgagg  ttagcatacg  tatcaaggac  agtaactacc  atggctcccg  aagttttgcc     120
aaaacctcgg  atgcgtggcc  ttctggcagg  cgtctgcgaa  atcat  atg  gct  gta  gca      177
                               Met Ala Val Ala
                               -10

```

ttc gtg cta tcc ctg ggg gtt gca gct ttg tat aag ttt cgt gtg gct 225
Phe Val Leu Ser Leu Gly Val Ala Ala Leu Tyr Lys Phe Arg Val Ala
-5 1 5

gat caa aga aag aag gca tac gca gat ttc tac aga aac tac gat gtc 273
Asp Gln Arg Lys Lys Ala Tyr Ala Asp Phe Tyr Arg Asn Tyr Asp Val
10 15 20

atg aaa gat ttt gag gag atg agg aag gct ggt atc ttt cag agt gta 321
Met Lys Asp Phe Glu Glu Met Arg Lys Ala Gly Ile Phe Gln Ser Val
25 30 35

aag taatcttgga atataaagaa tttcttcagg ttgaattacc tagaagtttg 374
Lys
40

t	c	a	c	t	g	a	c	t	t
g	t	g	t	t	c	c	t	g	a
a	c	t	a	t	g	a	c	a	c
a	t	g	a	a	t	a	t	g	t
g	g	g	c	t	a	a	g	a	a
a	t	a	g	t	t	c	c	t	c
:	:	:	:	:	:	:	:	:	:434
t	t	g	a	t	a	a	a	t	a
a	a	c	a	a	t	t	a	a	c
a	a	a	t	a	c	t	t	t	t
:	:	:	:	:	:	:	:	:	:466
g	:	:	:	:	:	:	:	:	:

<210> 390

<211> 522

<212> DNA

<213> Homo sapiens

 $\langle 220 \rangle$

<221> CDS

<222> 94..417

$\langle 220 \rangle$

```
<221> sig_peptide
```

<222> 94..213

<223> Von Heijne matrix

score 4

seq VTAPSMVAPVTFA/SI

<400> 390

```

ctgtaatctc tgcgaaagct tccaggttcg ctcatcgtag gtcactggaa aaaattgtag      60
cattgctaaa gagtatttca gagaaaattt tgc atg caa aaa ttt agg aag atg      114
                Met Gln Lys Phe Arg Lys Met
                -40                                -35

```

agt gaa act cat cat tct gta atc tct gtg aaa gct tcc agt ccc tgg 162
 Ser Glu Thr His His Ser Val Ile Ser Val Lys Ala Ser Ser Pro Trp
 -30 -25 -20

cta tct tct tca gtg act gct cca tcc atg gta gcc cca gtc act ttt 210
Leu Ser Ser Ser Val Thr Ala Pro Ser Met Val Ala Pro Val Thr Phe
-15 -10 -5

gca tct att gta gaa gaa gaa cta caa caa gaa gca gct ctt att aga 258
Ala Ser Ile Val Glu Glu Glu Leu Gln Gln Glu Ala Ala Leu Ile Arg
1 5 10 15

agt cga gaa aaa ccg ttg gct ctg att cag att gag gag cat gcc ata 306
 Ser Arg Glu Lys Pro Leu Ala Leu Ile Gln Ile Glu Glu His Ala Ile
 20 25 30

caa gat tta ttg gtt ttc tat gag gca ttt ggc aac cct gaa gag ttt 354
Gln Asp Leu Leu Val Phe Tyr Glu Ala Phe Gly Asn Pro Glu Glu Phe

224

gtc att gtt gaa agg aca ccg cag gga cca ctg gca gta cct atg tgg 402
 Val Ile Val Glu Arg Thr Pro Gln Gly Pro Leu Ala Val Pro Met Trp
 50 55 60
 aat aag cat gga tgc tagttcactg tggagttgag atgcatttta cataattatg 457
 Asn Lys His Gly Cys
 65
 agtttggttca tataaagaaa gctgtggaaa agagtcttag agatttgaat atcgtctaaa 517
 tagat 522

<210> 391
 <211> 409
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 9..389

<220>
 <221> sig_peptide
 <222> 9..149
 <223> Von Heijne matrix
 score 3.90000009536743
 seq LSSDLQQVGGASA/RI

<220>
 <221> misc_feature
 <222> 8,403
 <223> n=a, g, c or t

<400> 391
 tgggagtn atg ttc acg cct mtg ama gtg aaa tam gcg tac tac gac act 50
 Met Phe Thr Pro Xaa Xaa Val Lys Xaa Ala Tyr Tyr Asp Thr
 -45 -40 -35
 gaa cgc atc gga gtt gac mtg atc atg aag acc tgm ttt agc ccc aac 98
 Glu Arg Ile Gly Val Asp Xaa Ile Met Lys Thr Xaa Phe Ser Pro Asn
 -30 -25 -20
 aga gtg att gga ctc tca agt gac ttg cag caa gta gga ggg gca tca 146
 Arg Val Ile Gly Leu Ser Ser Asp Leu Gln Gln Val Gly Gly Ala Ser
 -15 -10 -5
 gct cgc atc mag gat gcc ctg agt aca gtg ttg caa tat gca gag gat 194
 Ala Arg Ile Xaa Asp Ala Leu Ser Thr Val Leu Gln Tyr Ala Glu Asp
 1 5 10 15
 gta ctg tct gga aag gtg tca gct gac aat act gtg ggc cgc ttc ctg 242
 Val Leu Ser Gly Lys Val Ser Ala Asp Asn Thr Val Gly Arg Phe Leu
 20 25 30
 atg agc ctg gtt aac caa gta ccg aaa ata gtt ccc gat gac ttt gag 290
 Met Ser Leu Val Asn Gln Val Pro Lys Ile Val Pro Asp Asp Phe Glu
 35 40 45
 acc atg ctc aac agc aac atc aat gac mtt ttg atg gtg acc tac mtg 338
 Thr Met Leu Asn Ser Asn Ile Asn Asp Xaa Leu Met Val Thr Tyr Xaa
 50 55 60
 gcc aac ctc aca cag tca cag att gca ctc aat gaa aaa ctt gta aac 386
 Ala Asn Leu Thr Gln Ser Gln Ile Ala Leu Asn Glu Lys Leu Val Asn
 65 70 75
 ctg tgaatggacc ccangcagta 409
 Leu
 80

<210> 392
 <211> 431
 <212> DNA
 <213> Homo sapiens

226

```

gat cgt gaa ttg gct gac atc cat gaa gat gcc aaa aca tgt ttg gta      285
Asp Arg Glu Leu Ala Asp Ile His Glu Asp Ala Lys Thr Cys Leu Val
      -20                      -15                      -10
cta tgt tcc aga gtg ctt tct gtc att tca gtc aag gaa ata aag aca      333
Leu Cys Ser Arg Val Leu Ser Val Ile Ser Val Lys Glu Ile Lys Thr
      -5                      1                      5
cag ctg agt tta gga aga cat cca att att tca aat tgg ttt gat tac      381
Gln Leu Ser Leu Gly Arg His Pro Ile Ile Ser Asn Trp Phe Asp Tyr
      10                      15                      20
att cct tca aca aga tac aaa gat cca tgt gaa cta tta cat ctt tgc      429
Ile Pro Ser Thr Arg Tyr Lys Asp Pro Cys Glu Leu Leu His Leu Cys
      30                      35                      40
aga cta acc atc agg aat caa cta tta acc aac aat atg ctc cca gat      477
Arg Leu Thr Ile Arg Asn Gln Leu Leu Thr Asn Asn Met Leu Pro Asp
      45                      50                      55
gga ata ttt tca ctt cta att cct gct cgt cta caa aac tat ctg aat      525
Gly Ile Phe Ser Leu Leu Ile Pro Ala Arg Leu Gln Asn Tyr Leu Asn
      60                      65                      70
tta gaa atc taacatacgt cagtgtc      551
Leu Glu Ile
      75

```

<210> 394

<211> 529

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 78..281

<220>

<221> sig_peptide

<222> 78..218

<223> Von Heijne matrix

score 3.70000004768372

seq PLLSLHSRGGSSS/ES

<400> 394

```

ttttttgcga acggcgagca gcggcgcgcg cgagagagacg agcggagggtt ttcctggttt      60
cggaccccag cggccgg atg gtg aaa tcc tcc ctg cag cgg atc ctc aat      110
      Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn
      -45                      -40
agc cac tgc ttc gcc aga gag aag gaa ggg gat aaa ccc agc gcc acc      158
Ser His Cys Phe Ala Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr
      -35                      -30                      -25
atc cac gcc agc cgc acc atg ccg ctc cta agc ctg cac agc cgc ggc      206
Ile His Ala Ser Arg Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly
      -20                      -15                      -10                      -5
ggc agc agc agt gag agt tcc agg gtc tcc ctc cac tgc tgt agt aac      254
Gly Ser Ser Ser Glu Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn
      1                      5                      10
ccg ggt ccg ggg cct cgg tgg tgc tcc tgatgccct caccaccccc      301
Pro Gly Pro Gly Pro Arg Trp Cys Ser
      15                      20
tgaagatccc aggtggggcga gggaatagtc agagggatca caatctttca gctaacttat      361
tctactccga tgatcggctg aatgtaacag aggaactaac gtccaacgac aagacgagga      421
ttctcaacgt ccagtccagg ctcacagacg ccaaacgcat taactggcga acagtgtcta      481
gtggcgccas tctacatcg agatcccggg cggcgcgctg ccgagggg      529

```

<210> 395

<211> 491

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 80..283

<220>

<221> sig_peptide

<222> 80..220

<223> Von Heijne matrix

score 3.70000004768372

seq PLLSLHSRGGSSS/ES

<400> 395

```

aktttgcgaa cggcgagcag cggcggcggc gcgagagac gcagcggagg ttttcctggt      60
ttcggacccc agcggccgg atg gtg aaa tcc tcc ctg cag cgg atc ctc aat      112
                Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn
                -45                                -40
agc cac tgc ttc gcc aga gag aag gaa ggg gat aaa ccc agc gcc acc      160
Ser His Cys Phe Ala Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr
-35                                -30                                -25
atc cac gcc agc cgc acc atg ccg ctc cta agc ctg cac agc cgc ggc      208
Ile His Ala Ser Arg Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly
-20                                -15                                -10                                -5
ggc agc agc agt gag agt tcc agg gtc tcc ctc cac tgs tgt agt aac      256
Gly Ser Ser Ser Glu Ser Ser Arg Val Ser Leu His Xaa Cys Ser Asn
                1                                5                                10
ccg ggt ccg ggg cct cgg tgg tgc tcc tgatgccct caccacccc      303
Pro Gly Pro Gly Pro Arg Trp Cys Ser
15                                20
tgaagatccc aggtgggcga gggaatagtc agagggatca caatctttca gctaacttat      363
tctactccga tgatcggctg aatgtaacag aggaactaac gtscaacgac aagacgagga      423
ttctcaacgt tcagtccagg ctacacaggaa cgccaaacgc attaactgcc gaacagtgct      483
gagtggcg                                491

```

<210> 396

<211> 503

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 9..395

<220>

<221> sig_peptide

<222> 9..56

<223> Von Heijne matrix

score 3.59999990463257

seq LLLLLRALRRGPG/PG

<400> 396

```

aggccaac atg gcc gtg ctg ctg ctg ctg ctc cgt gcc ctc cgc cgg ggt      50
                Met Ala Val Leu Leu Leu Arg Ala Leu Arg Arg Gly
                -15                                -10                                -5
cca ggc ccg ggt cct cgg ccg ctg tgg ggc cca ggc ccg gcc tgg agt      98
Pro Gly Pro Gly Pro Arg Pro Leu Trp Gly Pro Gly Pro Ala Trp Ser
1                                5                                10
cca ggg ttc ccc gcc agg ccc ggg agg ggg cgg ccg tac atg gcc agc      146
Pro Gly Phe Pro Ala Arg Pro Gly Arg Gly Arg Pro Tyr Met Ala Ser
15                                20                                25                                30
agg cct ccg ggg gac ctc gcc gag gct gga ggc cga gct ctg cag agc      194
Arg Pro Pro Gly Asp Leu Ala Glu Ala Gly Gly Arg Ala Leu Gln Ser

```

228

```

          35          40          45
tta caa ttg aga ctg cta acc cct acc ttt gaa ggg atc aac gga ttg      242
Leu Gln Leu Arg Leu Leu Thr Pro Thr Phe Glu Gly Ile Asn Gly Leu
          50          55          60
ttg ttg aaa caa cat tta gtt cag aat cca gtc aga ctc tgg caa ctt      290
Leu Leu Lys Gln His Leu Val Gln Asn Pro Val Arg Leu Trp Gln Leu
          65          70          75
tta ggt ggt act ttc tat ttt aac acc tca agg ttg aag cag aag aat      338
Leu Gly Gly Thr Phe Tyr Phe Asn Thr Ser Arg Leu Lys Gln Lys Asn
          80          85          90
aag gag aag gat aag tcg aag ggg aag gcg cct gaa gag gac gaa ggt      386
Lys Glu Lys Asp Lys Ser Lys Gly Lys Ala Pro Glu Glu Asp Glu Gly
          95          100          105          110
ata ttc atc tgatgttctt cagtcagtag ctgcctctgg atgtctttac      435
Ile Phe Ile
atttctgttt wccttttagc aaggtgaaac cagtcctggac aatggggaga tgggccgggt      495
gcagtggc      503

```

<210> 397
 <211> 517
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 93..419

<220>
 <221> sig_peptide
 <222> 93..191
 <223> Von Heijne matrix
 score 3.59999990463257
 seq LAVFQMLKSMCAG/QR

```

<400> 397
gaaaaaaagc gaaggccggc cgggcgggga agggaaatgg cgaggcagga gtgcggggga      60
gggastggtc cttagctgaa tgcgcctgcg tt atg gcg gcc tcc ggc gcc cca      113
                               Met Ala Ala Ser Gly Ala Pro
                               -30
agg atc ctg gtg gac ctg ctg aag ctg aac gtg gcc ccc ctc gcc gtc      161
Arg Ile Leu Val Asp Leu Leu Lys Leu Asn Val Ala Pro Leu Ala Val
-25          -20          -15
ttc cag atg ctc aag tcc atg tgt gcc ggg cag agg cta gcg agc gag      209
Phe Gln Met Leu Lys Ser Met Cys Ala Gly Gln Arg Leu Ala Ser Glu
-10          -5          1          5
ccc cag gac cct gcg gcc gtg tct ctg ccc acg tcg agc gtg ccc grg      257
Pro Gln Asp Pro Ala Ala Val Ser Leu Pro Thr Ser Ser Val Pro Xaa
          10          15          20
acc cga ggg aga aac aaa ggc agc gct gcc ctc ggg gga gca ttg gcc      305
Thr Arg Gly Arg Asn Lys Gly Ser Ala Ala Leu Gly Gly Ala Leu Ala
          25          30          35
ctg gcg gaa csm rrc agc cgc gaa gga tcc agc cag agg atg cca cgc      353
Leu Ala Glu Xaa Xaa Ser Arg Glu Gly Ser Ser Gln Arg Met Pro Arg
          40          45          50
cag ccc agc gct acc agg ctg ccc aag ggg ggc ggg cct ggg aag agc      401
Gln Pro Ser Ala Thr Arg Leu Pro Lys Gly Gly Gly Pro Gly Lys Ser
          55          60          65          70
cct aca cgg ggc agc acc taggatgggg cagagacttg ttgcatcttt      449
Pro Thr Arg Gly Ser Thr
          75
gtccccagca aaggctacat gttacctcct tcaattgata ataaaccttt ctgagatgag      509
agggtcca      517

```

```
<210> 398
<211> 507
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 78..386
```

```
<220>
<221> sig_peptide
<222> 78..185
<223> Von Heijne matrix
score 3.59999990463257
seq RALSTFLFGSIRG/AA
```

[illegible]

```
<210> 399
<211> 493
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 58..315
```

```
<220>
<221> sig_peptide
<222> 58..165
<223> Von Heijne matrix
score 3.59999990463257
seq SVSLSGFTVGTLS/ET
```

```

<400> 399
ctgtgggaga ggccaggatg ctgcattagg cacaggataa cctgggaacc caggcac      57
atg ggt cct gct ctc cga agt ctg caa gtc aag aag gga aca gag cac      105
Met Gly Pro Ala Leu Arg Ser Leu Gln Val Lys Lys Gly Thr Glu His
    -35                -30                -25

```

230

```

gcc gac cct ctc cct ttc ccc tct gtc tct ctt agt ggc ttt aca gtg      153
Ala Asp Pro Leu Pro Phe Pro Ser Val Ser Leu Ser Gly Phe Thr Val
-20                               -15                               -10                               -5
ggt acc ctg tca gaa acc agc act ggg ggc cct gcc acc ccc aca tgg      201
Gly Thr Leu Ser Glu Thr Ser Thr Gly Gly Pro Ala Thr Pro Thr Trp
                               1                               5                               10
aag gag tgt cct atc tgt aag gag cgc ttt cct gct gag agt gac aag      249
Lys Glu Cys Pro Ile Cys Lys Glu Arg Phe Pro Ala Glu Ser Asp Lys
                               15                               20                               25
gat gcc ctg gag gac cac atg gat gga cac ttc ttt ttc agc acc cag      297
Asp Ala Leu Glu Asp His Met Asp Gly His Phe Phe Ser Thr Gln
                               30                               35                               40
gac ccc ttc acc ttt gag tgatcttact ccctcgtaca tgcacaaata      345
Asp Pro Phe Thr Phe Glu
45                               50
cacactcatg cacacacaca ctacacacaca tgcatacact taggtttcat gccatttttc      405
tatcacactg ggctccatga tattctgttc cctaagaact gcttctgtgt gccctgtttt      465
catccaaga tttctcactt catcctct      493

```

<210> 400

<211> 769

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 506..661

<220>

<221> sig_peptide

<222> 506..595

<223> Von Heijne matrix

score 3.5

seq VKLLLMVSSATT/ER

<400> 400

```

cttggtcttt aaagccaggt tgagggttat gtgttttttat tcttttawat ctccttttca      60
cttaggcttg gcaactagtt cattgaaaca gagccaacaa gaatgagtga gtcactcagt      120
gatatgaagc aagctgtggt catttatcta ctaccttcca tacctacttc ctaggtttga      180
taagtagtgt ttgtagtgtg gttgtcagta actttttctca gottacataa gtttatgtgg      240
aattcattta caaatgtgtg attattacga tagatgttat tttataaatt tcaaggttct      300
ttaatacagc atggacatgt ttctaaatca ataaatatag cttamccaca gactttttat      360
gggatgtaaa gtatgtatta caacaggtaa atccatttct ctgttttggg aaagcagggt      420
aatgttttgt ttgttttgtt gctgcatata atgttgcaat atatatatat kctaagtcag      480
ttgggtctcct aatytctacag gataa atg tct agc tgt ggc att gtt gga tca      532
                               Met Ser Ser Cys Gly Ile Val Gly Ser
                               -30                               -25
tca gtt tcg ttt cag tta gat gct gtg aaa ttg ctc ttg aaa atg gtg      580
Ser Val Ser Phe Gln Leu Asp Ala Val Lys Leu Leu Leu Lys Met Val
-20                               -15                               -10
tcc tct gcc acc aca gaa cgg tgt tgt aat gga agt gcc aat ttt cat      628
Ser Ser Ala Thr Thr Glu Arg Cys Cys Asn Gly Ser Ala Asn Phe His
-5                               1                               5                               10
aaa aac ttg tgt gca aca ggt att aag aat ttt tgaggtcggg cgtagtagct      681
Lys Asn Leu Cys Ala Thr Gly Ile Lys Asn Phe
                               15                               20
cacgcctgta atcccagcac tttgggaggc ctaggcagggt ggatcacctg aggtcaggag      741
ttccagacca gcctgaccaa catggtga      769

```

<210> 401

<211> 646

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 76..621

<400> 401

```

attacgagat tggcttggat tctgtcggat ggacttgggg ctagctgcgg cggggctgga      60
ggaggccaga taacc atg tca gcc aca gtt gta gat gca gtt aat gct gca      111
          Met Ser Ala Thr Val Val Asp Ala Val Asn Ala Ala
          1          5          10
ccc cta tcg ggg tcc aaa gaa atg agt ttg gaa gaa cca aag aag atg      159
Pro Leu Ser Gly Ser Lys Glu Met Ser Leu Glu Glu Pro Lys Lys Met
          15          20          25
acc aga gag gac tgg aga aag aag aag gag cta gaa gaa cag cga aaa      207
Thr Arg Glu Asp Trp Arg Lys Lys Lys Glu Leu Glu Glu Gln Arg Lys
          30          35          40
ttg ggc aat gct cct gca gaa gtt gat gaa gaa gga aaa gac atc aac      255
Leu Gly Asn Ala Pro Ala Glu Val Asp Glu Glu Gly Lys Asp Ile Asn
          45          50          55          60
ccc cat att cct cag tat att tct tca gtg cca tgg tat att gat cct      303
Pro His Ile Pro Gln Tyr Ile Ser Ser Val Pro Trp Tyr Ile Asp Pro
          65          70          75
tca aaa aga cct act tta aaa cac cag aga cca caa cca gaa aaa caa      351
Ser Lys Arg Pro Thr Leu Lys His Gln Arg Pro Gln Pro Glu Lys Gln
          80          85          90
aag cag ttc agc tca tct gga gaa tgg tac aag agg ggt gta aaa gag      399
Lys Gln Phe Ser Ser Ser Gly Glu Trp Tyr Lys Arg Gly Val Lys Glu
          95          100          105
aat tcc ata att act aag tac cgc aaa gga gca tgt gaa aat tgt ggg      447
Asn Ser Ile Ile Thr Lys Tyr Arg Lys Gly Ala Cys Glu Asn Cys Gly
          110          115          120
gcc atg aca cac aaa aag aaa gac tgc ttt gag aga cct agg cga gtt      495
Ala Met Thr His Lys Lys Lys Asp Cys Phe Glu Arg Pro Arg Arg Val
          125          130          135          140
gga gcc aaa ttt aca ggt act aat ata gct cca gat gaa cat gtc cag      543
Gly Ala Lys Phe Thr Gly Thr Asn Ile Ala Pro Asp Glu His Val Gln
          145          150          155
cct caa ctg atg ttt gac tat gat ggg aag agg gat cgg tgg aat ggc      591
Pro Gln Leu Met Phe Asp Tyr Asp Gly Lys Arg Asp Arg Trp Asn Gly
          160          165          170
tac aat cca gaa gaa cac atg aaa att gtt tgaagagtat gcaaagtgat      641
Tyr Asn Pro Glu Glu His Met Lys Ile Val
          175          180
ttggc      646

```

<210> 402

<211> 484

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 29..469

<400> 402

```

acttggcgag tgagacgctg atgggagg atg gac gta ctg gtg tct gag tgc      52
          Met Asp Val Leu Val Ser Glu Cys
          1          5
tcc gcg cgg ctg ctg cag cag gaa gaa gag att aaa tct ctg act gct      100
Ser Ala Arg Leu Leu Gln Gln Glu Glu Glu Ile Lys Ser Leu Thr Ala
          10          15          20
gaa att gac cgg ttg aaa aac tgt ggc tgt tta gga gct tct cca aat      148
Glu Ile Asp Arg Leu Lys Asn Cys Gly Cys Leu Gly Ala Ser Pro Asn

```


232

25	30	35	40	
ttg gag cag tta caa gaa gaa aat tta aaa tta aag tat cga ctg aat				196
Leu Glu Gln Leu Gln Glu Glu Asn Leu Lys Leu Lys Tyr Arg Leu Asn				
	45	50	55	
att ctt cga aag agt ctt cag gca gaa agg aac aaa cca act aaa aat				244
Ile Leu Arg Lys Ser Leu Gln Ala Glu Arg Asn Lys Pro Thr Lys Asn				
	60	65	70	
atg att aac att att agc cgc cta caa gag gtc ttt ggt cat gca att				292
Met Ile Asn Ile Ile Ser Arg Leu Gln Glu Val Phe Gly His Ala Ile				
	75	80	85	
aag gct gca tat cca gat ttg gaa aat cct cct ctg cta gtg aca cca				340
Lys Ala Ala Tyr Pro Asp Leu Glu Asn Pro Pro Leu Leu Val Thr Pro				
	90	95	100	
agt cag cag gcc aag ttt ggg gac tat cag tgt aat agt gct atg ggt				388
Ser Gln Gln Ala Lys Phe Gly Asp Tyr Gln Cys Asn Ser Ala Met Gly				
105	110	115	120	
att tct cag gtg atg tat tgt cat gac tct tgg ctg ttt gat ttt ttt				436
Ile Ser Gln Val Met Tyr Cys His Asp Ser Trp Leu Phe Asp Phe Phe				
	125	130	135	
aag tat tat tat cat cat tgc cat tta cag aaa taatactatt acaag				484
Lys Tyr Tyr Tyr His His Cys His Leu Gln Lys				
	140	145		

<210> 403

<211> 345

<212> DNA

<213> Homo sapiens

<400> 403

ctttctccag	ctcaagcgctc	tggtatatttg	tgtagggaca	gcgcttttttc	cggttgaggc	60
gagcgtggat	ccagttcgcg	gcggggttg	ttgggtcaag	ttgctgctgc	tgctgcggct	120
gcgaatgctg	cttctctctc	tccttcagcg	aacagcctgg	gatcggtgg	tcgctgcaag	180
ctgagggctc	cgacgagccg	cccgtcccag	tcgcgggccc	gatccctgcc	ccaggcccgg	240
ttccccccgt	cgccgctgcc	gccttctcct	gcaggaacga	gttgcacatg	ttactgcttc	300
ctactaataa	atgctgacct	gatcaaatgg	agcccaaaar	aaaaa		345

<210> 404

<211> 432

<212> DNA

<213> Homo sapiens

<400> 404

accaggactc	caaaatggcg	tcagttgtac	cagtgaagga	caagaaactt	ctggagggtca	60
aactggggga	gctgccaagc	tggatcttga	tgcgggactt	cagtctagtg	gcatttttcgg	120
agcgtttcaa	agaggttact	accggtacta	caacaagtac	atcaatgtga	agaaggggag	180
catctcgggg	attaccatgg	tgctggcatg	ctacgtgctc	tttagctact	ccttttccta	240
caagcatctc	aagcacgagc	ggctccgcaa	ataccactga	agaggacaca	ctctgcaccc	300
ccccaccca	cgaccttggc	ccgagcccct	ccgtgagaac	acaatctcaa	tcgttgctga	360
atcctttcat	atcctaatag	gaattaacct	ccaaataaaa	catgactggt	acgyraaaaa	420
aaaaaaaaaa	ag					432

<210> 405

<211> 428

<212> DNA

<213> Homo sapiens

<400> 405

actccgcctt	ccacgtgcac	ccactgcctc	ttcccttctc	gcttgggaac	tctagtctcg	60
cctcgggttg	caatggaccc	caactgctcc	tgtgccgctg	caggtgtctc	ctgcacctgc	120
gccagctcct	gcaagtgcaa	agagtgcaaa	tgcacctcct	gcaagaagag	ctgctgctcc	180
tgctgccttg	tggtgtgtgc	caagtgtgcc	cagggctgca	tctgcaaagg	ggcatcggag	240
aagtgcagct	gctgcgcctg	atgtcgggac	agccctgctc	ccaagtacaa	atagagtgc	300
ccgtaaaatc	caggatTTTT	tgTTTTTtgc	tacaatcttg	acccctttgc	tacattcctt	360

233

tttttctgtg aaatatgtga ataataatta aacacttaga caaaaaaaaa aaaaaaaggt 420
caaaagct 428

<210> 406
<211> 391
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 196..390

<220>
<221> sig_peptide
<222> 196..279
<223> Von Heijne matrix
score 5.90000009536743
seq IILLLLKIWKSR/CA

<400> 406
ctcttcccg gctccagctcc gccgccagct ccagcctttg ctccccctcc caaagtcccc 60
tccccggagc ggagcgcacc tagggtcctt cttccgtccc cccagcccag ctaccggttc 120
agaccagcag cctcgggggg ccccccccg ccagcctgcc tccctcccg ctagccctgc 180
cagggttccc cagcc atg aat ctc ttc cga ttc ctg gga gac ctc tcc cac 231
Met Asn Leu Phe Arg Phe Leu Gly Asp Leu Ser His
-25 -20
ctc ctc gcc atc atc ttg cta ctg ctc aaa atc tgg aag tcc cgc tcg 279
Leu Leu Ala Ile Ile Leu Leu Lys Ile Trp Lys Ser Arg Ser
-15 -10 -5
tgc gcc gcc cac ccc cag ctt cct ctc tcc ttc tgt ctg tct gtc tgt 327
Cys Ala Ala His Pro Gln Leu Pro Leu Ser Phe Cys Leu Ser Val Cys
1 5 10 15
ctg tct gtc tct ctc tct ctc tct gtc tct ctc tct ctc tct ttc tct 375
Leu Ser Val Ser Leu Ser Leu Ser Val Ser Leu Ser Leu Ser Phe Ser
20 25 30
gtc tca aaa aaa aaa t 391
Val Ser Lys Lys Lys
35

<210> 407
<211> 500
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 81..500

<220>
<221> sig_peptide
<222> 81..200
<223> Von Heijne matrix
score 5.19999980926514
seq LLMRALHLLKSGC/SP

<400> 407
aaccccaccg cctaactctc ccaactctgc agtccgtgac cttcaggtgc ctttcaaggt 60
cagaaaaact cggcaagaat atg gtt tct agt ttt agg gtt tct gaa cta caa 113
Met Val Ser Ser Phe Arg Val Ser Glu Leu Gln
-40 -35 -30
gta tta cta ggc ttt gct gga cgg aat aaa agt gga cgc aag cat gac 161
Val Leu Leu Gly Phe Ala Gly Arg Asn Lys Ser Gly Arg Lys His Asp
-25 -20 -15

234

ctc	ctg	atg	agg	gcg	ctg	cat	tta	ttg	aag	agc	ggc	tgc	agc	cct	gcg	209
Leu	Leu	Met	Arg	Ala	Leu	His	Leu	Leu	Lys	Ser	Gly	Cys	Ser	Pro	Ala	
		-10						-5					1			
gtt	cag	att	aaa	atc	cga	gaa	ttg	tat	aga	cgc	cga	tat	cca	cga	act	257
Val	Gln	Ile	Lys	Ile	Arg	Glu	Leu	Tyr	Arg	Arg	Arg	Tyr	Pro	Arg	Thr	
	5					10				15						
ctt	gaa	gga	ctt	tct	gat	tta	tcc	aca	atc	aaa	tca	tcg	gtt	ttc	agt	305
Leu	Glu	Gly	Leu	Ser	Asp	Leu	Ser	Thr	Ile	Lys	Ser	Ser	Val	Phe	Ser	
20				25						30				35		
ttg	gat	ggg	ggc	tca	tca	cct	gta	gaa	cct	gac	ttg	gcc	gtg	gct	gga	353
Leu	Asp	Gly	Gly	Ser	Ser	Pro	Val	Glu	Pro	Asp	Leu	Ala	Val	Ala	Gly	
				40					45					50		
atc	cac	tcg	ttg	cct	tcc	act	tca	gtt	aca	cct	cac	tca	cca	tcc	tct	401
Ile	His	Ser	Leu	Pro	Ser	Thr	Ser	Val	Thr	Pro	His	Ser	Pro	Ser	Ser	
			55					60					65			
cct	gtt	ggg	tct	gtg	ctg	ctt	caa	gat	act	aag	ccc	aca	ttt	gag	atg	449
Pro	Val	Gly	Ser	Val	Leu	Leu	Gln	Asp	Thr	Lys	Pro	Thr	Phe	Glu	Met	
	70						75				80					
cag	cag	cca	tct	ccc	cca	att	cct	cct	gtc	cat	cct	gat	gtg	cag	tta	497
Gln	Gln	Pro	Ser	Pro	Pro	Ile	Pro	Pro	Val	His	Pro	Asp	Val	Gln	Leu	
	85					90					95					
aaa																500
Lys																
100																

<210> 408

<211> 497

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 35..496

<400> 408

tccta	atcga	aawct	tcttct	gaggat	ccgg	caag	atg	gca	gaa	gta	gag	cag	aag			55
								Met	Ala	Glu	Val	Glu	Gln	Lys		
								1				5				
aag	aag	cgg	acc	ttc	cg	aag	ttc	acc	tac	cg	ggc	gtg	gac	ctc	gac	103
Lys	Lys	Arg	Thr	Phe	Arg	Lys	Phe	Thr	Tyr	Arg	Gly	Val	Asp	Leu	Asp	
		10					15				20					
cag	ctg	ctg	gac	atg	tcc	tac	gag	cag	ctg	atg	cag	ctg	tac	agt	gcg	151
Gln	Leu	Leu	Asp	Met	Ser	Tyr	Glu	Gln	Leu	Met	Gln	Leu	Tyr	Ser	Ala	
	25					30					35					
cg	cag	gcg	gcg	gct	gaa	ccg	ggg	cct	gcg	gcg	gaa	gca	gca	ctc	cct	199
Arg	Gln	Ala	Ala	Ala	Glu	Pro	Gly	Pro	Ala	Ala	Glu	Ala	Ala	Leu	Pro	
	40				45					50					55	
gct	gaa	gcg	cct	gcg	caa	ggc	caa	gaa	gga	ggc	gcc	gcc	cat	gga	gaa	247
Ala	Glu	Ala	Pro	Ala	Gln	Gly	Gln	Glu	Gly	Gly	Ala	Ala	His	Gly	Glu	
			60						65					70		
gcc	gga	agt	ggg	gaa	gac	gca	cct	gcg	gga	cat	gat	cat	cct	acc	cga	295
Ala	Gly	Ser	Gly	Glu	Asp	Ala	Pro	Ala	Gly	His	Asp	His	Pro	Thr	Arg	
		75						80					85			
gat	ggg	ggg	cag	cat	ggg	ggg	cgt	cta	caa	cg	caa	gac	ctt	caa	cca	343
Asp	Gly	Gly	Gln	His	Gly	Gly	Arg	Leu	Gln	Arg	Gln	Asp	Leu	Gln	Pro	
	90					95						100				
ggg	gga	gat	caa	gcc	cga	gat	gat	cg	cca	cta	cct	ggg	cga	gtt	ctc	391
Gly	Gly	Asp	Gln	Ala	Arg	Asp	Asp	Arg	Pro	Leu	Pro	Gly	Arg	Val	Leu	
	105					110					115					
cat	cac	cta	caa	gcc	cgt	aaa	gca	tgg	ccg	gcc	cg	cat	cg	ggc	mmc	439
His	His	Leu	Gln	Ala	Arg	Lys	Ala	Trp	Pro	Ala	Arg	His	Arg	Gly	Xaa	
	120				125				130						135	
cma	mtc	ctc	csg	ctt	cat	ccm	tct	caa	gta	atg	gct	cag	cta	ata	aag	487

235

Xaa Xaa Leu Xaa Leu His Pro Ser Gln Val Met Ala Gln Leu Ile Lys
 140 145 150
 gcg cac atg a
 Ala His Met

497

<210> 409
 <211> 304
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 110..304

<220>
 <221> misc_feature
 <222> 296..297
 <223> n=a, g, c or t

<400> 409
 acgtacggca gggcggggtc ggcccgcagt ggccagctgg gagctacgag cgcggasttg 60
 cgcagaagac ccccatcagg gtgcgggggtg cagttgcggc tccagggcc atg gcg gag 118
 Met Ala Glu
 1
 gag cag ggc cgg gaa cgg gac tcg gtt ccc aag ccg tcg gtg ctg ttc 166
 Glu Gln Gly Arg Glu Arg Asp Ser Val Pro Lys Pro Ser Val Leu Phe
 5 10 15
 ctc cac cca gac ctg ggc gtg ggc ggc gct gag cgg ctg gtg ctc ttg 214
 Leu His Pro Asp Leu Gly Val Gly Gly Ala Glu Arg Leu Val Leu Leu
 20 25 30 35
 cct ttt ccc act gag aga agg ctg ctc ttt tgt act gcc ccc cgc tca 262
 Pro Phe Pro Thr Glu Arg Arg Leu Leu Phe Cys Thr Ala Pro Arg Ser
 40 45 50
 tta aac agc ctc ccc cta aaa aaa aaa aaa att nna aaa agc 304
 Leu Asn Ser Leu Pro Leu Lys Lys Lys Lys Ile Xaa Lys Ser
 55 60 65

<210> 410
 <211> 553
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 155..553

<400> 410
 aatagtgagc ggcagacaga ggggtggcca gtttgttggtt ttacctaaaa ggtgttttcc 60
 acatggagtg ggactgagag gcacaggtgg ctgaggggac ccgcctggga cgtgaggcgc 120
 aggtcctcgg ccccttcac tgtcagctgc agca atg gaa tac gtg ctg gaa gtg 175
 Met Glu Tyr Val Leu Glu Val
 1 5
 aag aac tct ccg cgg cac ctc ctg aag caa ttc aca gtg tgt gac gtt 223
 Lys Asn Ser Pro Arg His Leu Lys Gln Phe Thr Val Cys Asp Val
 10 15 20
 cct ctg tat gac att tgt gac tac aac gtc tcc aga gac cga tgc cag 271
 Pro Leu Tyr Asp Ile Cys Asp Tyr Asn Val Ser Arg Asp Arg Cys Gln
 25 30 35
 gag ctc ggg tgc tgc ttc tac gaa ggc gtc tgc tac aag aaa gcg gtt 319
 Glu Leu Gly Cys Cys Phe Tyr Glu Gly Val Cys Tyr Lys Lys Ala Val
 40 45 50 55
 ccc att tac atc cac gtg ttc tct gcc ttg att gtg atc ata gct ggg 367
 Pro Ile Tyr Ile His Val Phe Ser Ala Leu Ile Val Ile Ile Ala Gly

236

60	65	70	
gcc ttc gtc atc acc atc atc tac aga gtc att cag gag agc agg aaa			415
Ala Phe Val Ile Thr Ile Ile Tyr Arg Val Ile Gln Glu Ser Arg Lys			
75	80	85	
gaa aag gcc atc cct gtg tat gtc gcg ctg cca cag aag tcc agc gaa			463
Glu Lys Ala Ile Pro Val Tyr Val Ala Leu Pro Gln Lys Ser Ser Glu			
90	95	100	
aag gcg gag ttg gcc tca tcc agc agc aag tta ggg ctg aag ctg cga			511
Lys Ala Glu Leu Ala Ser Ser Ser Ser Lys Leu Gly Leu Lys Leu Arg			
105	110	115	
gtc ctg ggc ctc aaa gtg ctg ggc ctc gat gag agt gac gag			553
Val Leu Gly Leu Lys Val Leu Gly Leu Asp Glu Ser Asp Glu			
120	125	130	

<210> 411
 <211> 470
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 23..469

<400> 411

gggttcccgt tccccgcgga sc atg cgg tac aac gag aag gag ctg cag gct	52
Met Arg Tyr Asn Glu Lys Glu Leu Gln Ala	
1	5
ctg tcc cgg cag ccg gcc gag atg gcg gcc gag ctg ggc atg agg ggc	100
Leu Ser Arg Gln Pro Ala Glu Met Ala Ala Glu Leu Gly Met Arg Gly	
15	20
ccc aag aag ggc agc gtg ctg aag cgg cgg ctg gtg aag ctg gtg gtg	148
Pro Lys Lys Gly Ser Val Leu Lys Arg Arg Leu Val Lys Leu Val Val	
30	35
aat ttc ctc ttc tac ttt cgg aca gac gag gcc gag ccc gtc gga gcc	196
Asn Phe Leu Phe Tyr Phe Arg Thr Asp Glu Ala Glu Pro Val Gly Ala	
45	50
ctg ctg ctg gag cgc tgc aga gtc gtc cgg gaa gag ccc ggc acc ttc	244
Leu Leu Leu Glu Arg Cys Arg Val Val Arg Glu Glu Pro Gly Thr Phe	
60	65
tcc atc agc ttc att gag gac cct gag agg aag tat cac ttt gag tgc	292
Ser Ile Ser Phe Ile Glu Asp Pro Glu Arg Lys Tyr His Phe Glu Cys	
75	80
agc agc gag gag cag tgt cag gag tgg atg gag gct ctg cgt cgg gcc	340
Ser Ser Glu Glu Gln Cys Gln Glu Trp Met Glu Ala Leu Arg Arg Ala	
95	100
agc tac gag ttc atg cgg aga agc ctc atc ttc tac agg aac gaa atc	388
Ser Tyr Glu Phe Met Arg Arg Ser Leu Ile Phe Tyr Arg Asn Glu Ile	
110	115
cgg aag gtg acg ggc aag gac ccc ctg gaa cag ttc ggc ata tcc gag	436
Arg Lys Val Thr Gly Lys Asp Pro Leu Glu Gln Phe Gly Ile Ser Glu	
125	130
gag gcc agg ttc cag ctg agt ggc ttg cag gcg t	470
Glu Ala Arg Phe Gln Leu Ser Gly Leu Gln Ala	
140	145

<210> 412
 <211> 486
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 211..486

<400> 412

```

agattccttaa ctacgtgtct tacagcaaag ttgcccatgg gaacaaaggc tgtaggggtgt      60
cattgatctt gtaatacaac acatatTTga tgggtgaatc ttattagcac tcctttgaaa      120
atgttccagt atggagtcag ttgctgataa agaggaagtt cacagctagg aatccgaagt      180
taagatttct caattgccgc ctgggcactc atg cct gta gtc cca gca ctt ggg      234
                               Met Pro Val Val Pro Ala Leu Gly
                               1           5
agg cca agg tgg gca gat cac ctg agg tgc gga gtt cgg gac cag cct      282
Arg Pro Arg Trp Ala Asp His Leu Arg Ser Gly Val Arg Asp Gln Pro
    10           15           20
ggc caa cct ggt gaa gcc ccc cca tct cta cta aaa ata caa aaa ttg      330
Gly Gln Pro Gly Glu Ala Pro Pro Ser Leu Leu Lys Ile Gln Lys Leu
    25           30           35           40
gcc ggg tat ggt ggt ggc tgc ctg tgg tcc cag cta ctc ggg agg ctg      378
Ala Gly Tyr Gly Gly Gly Cys Leu Trp Ser Gln Leu Leu Gly Arg Leu
           45           50           55
agg cgg gag aat cac ttg agc ccg gga ggc gga ggt tgc agt gag ccg      426
Arg Arg Glu Asn His Leu Ser Pro Gly Gly Gly Cys Ser Glu Pro
           60           65           70
aga ttg tgc cac tgc act cca gcc tgg gtg aca gag caa gac tcc atc      474
Arg Leu Cys His Cys Thr Pro Ala Trp Val Thr Glu Gln Asp Ser Ile
           75           80           85
tca aaa ata gaa      486
Ser Lys Ile Glu
    90

```

<210> 413

<211> 335

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 73..252

<220>

<221> sig_peptide

<222> 73..135

<223> Von Heijne matrix

score 4.30000019073486

seq AGVSCTCAXPASA/KS

<400> 413

```

actccgcctt ccacgtgcac ccactgcctc ttcccttctc gcttgggaac tctagtctcg      60
cctcggggttg ca atg gac ccc aac tgc tcc tgt gcc gct ggt gtc tcc tgc      111
                               Met Asp Pro Asn Cys Ser Cys Ala Ala Gly Val Ser Cys
                               -20           -15           -10
acc tgc gcc ast cct gca agt gca aag agt gca aat gca cct cct gca      159
Thr Cys Ala Xaa Pro Ala Ser Ala Lys Ser Ala Asn Ala Pro Pro Ala
           -5           1           5
aga aga gct gct gct cct ggt gca ctg gcc cca ccc tgc tgg aca cct      207
Arg Arg Ala Ala Ala Pro Gly Ala Leu Ala Pro Pro Ser Trp Thr Pro
    10           15           20
gcc ctg ccc tgc cac ctg tct gtc tgt ccc aaa gaa gtt ctg gta      252
Ala Leu Pro Cys His Leu Ser Val Cys Pro Lys Glu Val Leu Val
    25           30           35
tgaacttgag gacacatgtc cagtgggagg tgagaccacc tctcaatatt caataaagct      312
gctgagaatc tagcctcgaa aaa      335

```

<210> 414

<211> 494

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 63..422

<220>

<221> sig_peptide

<222> 63..167

<223> Von Heijne matrix

score 4.09999990463257

seq LSQSLAMSPQTHS/QT

<400> 414

```

cacggatctc acaggtggag gtggggcgag cttctttcta gtgtggacca gaggtggggg      60
ca atg agt aaa cat ctt cac tgt aag gtt ttg gga ctg ccc ctc agg      107
  Met Ser Lys His Leu His Cys Lys Val Leu Gly Leu Pro Leu Arg
    -35          -30          -25
ctt gag gtc tcg gca tcg tgt ctc agt cag tcc ttg gcg atg tcc cct      155
Leu Glu Val Ser Ala Ser Cys Leu Ser Gln Ser Leu Ala Met Ser Pro
-20          -15          -10          -5
cag acg cac tca cag act tgc ata cgt aat tta gta aca tgc atc aac      203
Gln Thr His Ser Gln Thr Cys Ile Arg Asn Leu Val Thr Cys Ile Asn
          1          5          10
tat cct aga aca tct aca ggg tgc aaa gga acc acc act cag aga atc      251
Tyr Pro Arg Thr Ser Thr Gly Cys Lys Gly Thr Thr Thr Gln Arg Ile
          15          20          25
atg gag cca gtg gag tta gaa gtg gaa ggg aca gaa caa gac aat gct      299
Met Glu Pro Val Glu Leu Glu Val Glu Gly Thr Glu Gln Asp Asn Ala
          30          35          40
aaa acc tgt ggt tca cta gga agg ggg aat gaa aac acc atg ctc cga      347
Lys Thr Cys Gly Ser Leu Gly Arg Gly Asn Glu Asn Thr Met Leu Arg
          45          50          55          60
ggt gga ttc agc atg aac aca act gtg ggg caa gga att tcc aag caa      395
Gly Gly Phe Ser Met Asn Thr Thr Val Gly Gln Gly Ile Ser Lys Gln
          65          70          75
aca cac cac act agt acc act tct tcc taaacagaaa agggaaagag      442
Thr His His Thr Ser Thr Thr Ser Ser
          80          85
catccttcga cttccccctc cctccatcag caccgggttt tcctocaaaa aa      494

```

<210> 415

<211> 366

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 112..306

<220>

<221> sig_peptide

<222> 112..153

<223> Von Heijne matrix

score 3.70000004768372

seq LCWLKLVSIITVA/II

<400> 415

```

cctttctcgt tccccggcca tcttatcggc tgctgttggt tggggggccgt cccgctccta      60
aggcaggaag atggtggccg caaagaagac ggaaatctga aatagagtac t atg cta      117
                                Met Leu
tgt tgg cta aaa ctg gtg tcc atc act aca gtg gca ata ata ttg aac      165
Cys Trp Leu Lys Leu Val Ser Ile Thr Thr Val Ala Ile Ile Leu Asn

```

239

```

      -10      -5      1
tgg gca cag cat gcg gaa aat act aca gag tgt gca cac tgg cta tca      213
Trp Ala Gln His Ala Glu Asn Thr Thr Glu Cys Ala His Trp Leu Ser
5          10          15          20
ttg atc cag gtg act ctg aca tca tta gaa gca tgc cag aac aga ctg      261
Leu Ile Gln Val Thr Leu Thr Ser Leu Glu Ala Cys Gln Asn Arg Leu
          25          30          35
gtg aaa agt aaa cct ttt cac cta caa aat ttc acc tgc aaa cct      306
Val Lys Ser Lys Pro Phe His Leu Gln Asn Phe Thr Cys Lys Pro
          40          45          50
taaacctgca aaattttcct ttaataaaat ttgcttggttc taaaaaaaaa aaaaaaaaaag      366

```

<210> 416
 <211> 472
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 115..411

```

<400> 416
aagaaagggg tgaggcctaa gggacaatca ggatgttttt cagagagaag tgtggatgct      60
ggacaggaag aaccacagat accagatacg ggtactgttg taactctgtt ctcc atg      117
                                   Met
                                   1
aaa aaa aag gaa gaa aca aca ctt tca gag atg gag cct gtt gag cca      165
Lys Lys Lys Glu Glu Thr Thr Leu Ser Glu Met Glu Pro Val Glu Pro
          5          10          15
cag tac caa cta gtc aat gct gaa tcg act tct ccc ttt cta cat tgc      213
Gln Tyr Gln Leu Val Asn Ala Glu Ser Thr Ser Pro Phe Leu His Cys
          20          25          30
ctg aga gaa gtc att ggg gaa tac tct gta cac gaa ttt tca ctg ttg      261
Leu Arg Glu Val Ile Gly Glu Tyr Ser Val His Glu Phe Ser Leu Leu
          35          40          45
ggg aaa aca gag agt caa ggg att gga ttg tgg att gca ttg gtg gtt      309
Gly Lys Thr Glu Ser Gln Gly Ile Gly Leu Trp Ile Ala Leu Val Val
          50          55          60          65
ttc ctc agt ttc ctc atc ttc tcc aca agt ttc tac ata tcg aat gca      357
Phe Leu Ser Phe Leu Ile Phe Ser Thr Ser Phe Tyr Ile Ser Asn Ala
          70          75          80
gag cag ccc ttc ttc aaa gaa cct cct acg gaa gct gct aag gaa ctc      405
Glu Gln Pro Phe Phe Lys Glu Pro Pro Thr Glu Ala Ala Lys Glu Leu
          85          90          95
agt ctg tagctctgcg tggagccatg tgtaaacact gaactgagac ctgccacctc      461
Ser Leu
ctactaccta a      472

```

<210> 417
 <211> 532
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 125..508

```

<400> 417
agctrctctgg gtaaggccca agatggctgt cttcgccctta gtactcgtgt gaagttggcg      60
gggacgggttc ctgtcatctt cttgggctta tttggtgtgc tgttgaagg gggagactag      120
agaa atg gca ggg aac ctc tta tcc ggg gca ggt agg cgc ctg tgg gac      169
Met Ala Gly Asn Leu Leu Ser Gly Ala Gly Arg Arg Leu Trp Asp
          1          5          10          15

```


240

tgg	gtg	cct	ctg	gcg	tgc	aga	agc	ttc	tct	ctt	ggt	gtg	cct	aga	ttg	217
Trp	Val	Pro	Leu	Ala	Cys	Arg	Ser	Phe	Ser	Leu	Gly	Val	Pro	Arg	Leu	
			20					25						30		
atc	ggt	ata	agg	ctc	act	ctc	ccg	ccc	ccc	aaa	gtg	gtt	gat	cgt	tgg	265
Ile	Gly	Ile	Arg	Leu	Thr	Leu	Pro	Pro	Pro	Lys	Val	Val	Asp	Arg	Trp	
			35					40						45		
aac	gag	aaa	agg	gcc	atg	ttc	gga	gtg	tat	gac	aac	atc	ggg	atc	ctg	313
Asn	Glu	Lys	Arg	Ala	Met	Phe	Gly	Val	Tyr	Asp	Asn	Ile	Gly	Ile	Leu	
		50					55					60				
gga	aac	ttt	gaa	aag	cac	mcc	aaa	gaa	ctg	atc	agg	ggg	ccc	ata	tgg	361
Gly	Asn	Phe	Glu	Lys	His	Xaa	Lys	Glu	Leu	Ile	Arg	Gly	Pro	Ile	Trp	
	65				70					75						
ctt	cga	ggt	tgg	aaa	ggg	aat	gaa	ttg	caa	cgt	tgt	atc	cga	aag	agg	409
Leu	Arg	Gly	Trp	Lys	Gly	Asn	Glu	Leu	Gln	Arg	Cys	Ile	Arg	Lys	Arg	
80					85					90					95	
aaa	atg	gtt	gga	agt	aga	atg	ttc	gct	gat	gac	ctg	cac	aac	ctt	aat	457
Lys	Met	Val	Gly	Ser	Arg	Met	Phe	Ala	Asp	Asp	Leu	His	Asn	Leu	Asn	
			100						105					110		
aaa	cgc	atc	cgc	tat	ctc	tac	aaa	cac	ttt	aac	cga	cat	ggg	aag	ttt	505
Lys	Arg	Ile	Arg	Tyr	Leu	Tyr	Lys	His	Phe	Asn	Arg	His	Gly	Lys	Phe	
			115					120					125			
cga	tagaagagaa	agctgagaac	ttcg													532
Arg																

<210> 418

<211> 470

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 50..205

<400> 418

actctttcct	tggctgctca	ggtcataagg	gctgtgatct	agagacaaa	atg	ttc	ctt	58								
					Met	Phe	Leu									
					1											
aca	ctg	gca	gat	act	tgc	aaa	cta	agg	gga	atg	agc	ttc	ctt	ctg	aat	106
Thr	Leu	Ala	Asp	Thr	Cys	Lys	Leu	Arg	Gly	Met	Ser	Phe	Leu	Leu	Asn	
	5					10				15						
gtt	tat	gaa	gga	gag	gcc	act	gtg	tca	tct	gtc	tta	gag	cta	ttg	gaa	154
Val	Tyr	Glu	Gly	Glu	Ala	Thr	Val	Ser	Ser	Val	Leu	Glu	Leu	Leu	Glu	
20					25					30					35	
tcc	tgg	atc	att	gtg	gga	aat	gaa	aga	tac	ttc	gat	gga	atc	agc	agc	202
Ser	Trp	Ile	Ile	Val	Gly	Asn	Glu	Arg	Tyr	Phe	Asp	Gly	Ile	Ser	Ser	
			40					45					50			
cat	tgatccaatg	ccaactccaa	gactggaacg	tcgcaatgat	agttccaagg											255
His																
cggaattttg	acgtaattct	tttcgacaca	gttttacagg	tctggatgcc	cattttaatt											315
cttctgaaag	catgcctyct	ccttctggct	tcaggactcc	atctscagcc	tcttgatcta											375
aaaataatcc	ccaaacccaa	aaaaaattag	atactatttc	ctcaaaatta	ggatatttta											435
tcaaaaacat	cttaacataa	atacattatt	atcac													470

<210> 419

<211> 559

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 175..426

<400> 419

241

```

actcggcggg cgctgttgag ggagtcgggc cgcgactgtg gtcgttttta taccttcccg      60
cgcggacgcc ggcgtgcca acggaagggc gggtagggcg agacggagtt tcgtcatgtt      120
ggccaggccc atttgagatc tttgaagata tcctcaacgt gaggctctgc tgcc atg      177
                                   Met
                                   1
aag gtg aag att aag tgc tgg aac ggc gtg gcc act tgg ctc tgg gtg      225
Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp Val
                                   5
gcc aac gat gag aac tgt ggc atc tgc agg atg gca ttt aac gga tgc      273
Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly Cys
                                   20
tgc cct gac tgc aag gtg ccc ggc gac gac tgc ccg ctg gtg tgg ggc      321
Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp Gly
                                   35
cag tgc tcc cac tgc ttc cac atg cat tgc atc ctc aag tgg ctg cac      369
Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu His
                                   50
gca cag cag gtg cag cag cac tgc ccc atg tgc cgc cag gaa tgg aag      417
Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp Lys
                                   70
ttc aag gag tgaggccccga cctggctctc gctggagggg catcctgaga      466
Phe Lys Glu
ctccttcttc atgctggcgc cgatggctgc tggggacagc gcccctgagc tgcaacaagg      526
tgawacaag ggctggagct gcgtttgttt tgc      559

```

<210> 420

<211> 462

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 29..268

<400> 420

```

atccagcagc cccctgctcc ggcccagc atg gcg acc ccg acc cag acc ccc      52
                                   Met Ala Thr Pro Thr Gln Thr Pro
                                   1
aca aag gct cct gag gaa cct gac cca ttt tac tat gac tac aac acg      100
Thr Lys Ala Pro Glu Glu Pro Asp Pro Phe Tyr Tyr Asp Tyr Asn Thr
                                   10
gtg cag act gtg ggc atg act ctg gca acc atc ttg ttc ctg ctg ggt      148
Val Gln Thr Val Gly Met Thr Leu Ala Thr Ile Leu Phe Leu Leu Gly
                                   25
atc ctc atc gtc atc agc aag aag gtg aag tgc agg aag gcg gac tcc      196
Ile Leu Ile Val Ile Ser Lys Lys Val Lys Cys Arg Lys Ala Asp Ser
                                   45
agg tct gag agc cca acc tgc aaa tcc tgt aag tct gag ctt ccc tct      244
Arg Ser Glu Ser Pro Thr Cys Lys Ser Cys Lys Ser Glu Leu Pro Ser
                                   60
tca gcc cct ggt ggc ggc ggc gtg taacaccttc ccgaggaaac tccgctgccg      298
Ser Ala Pro Gly Gly Gly Gly Val
                                   75
accctgcctg agcgcgggag cctgaggacc gggtaggggc ggtggggacc cagccgcgcg      358
ccgggagcgc tccccgaat gagcgccca cccaccccaa ggctggagcc gctgcaccct      418
gctgtccctc tccaggcctt ggcaatgacg atcccccaaa gagc      462

```

<210> 421

<211> 547

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 99..479

<400> 421

```

tttcgaaggg aagggagccc cctataaaac agcctacagt ggacagtctg gtcggcagag      60
ccgcaggtca gtcgtgaaga gggagctcta ttgccacc atg agt ttc tcc ggc aag      116
                               Met Ser Phe Ser Gly Lys
                               1           5
tac caa ctg cag agc cag gaa aac ttt gaa gcc ttc atg aag gca atc      164
Tyr Gln Leu Gln Ser Gln Glu Asn Phe Glu Ala Phe Met Lys Ala Ile
           10           15           20
ggc ctg ccg gaa gag ctc atc cag aag ggg aag gat atc aag ggg gtg      212
Gly Leu Pro Glu Glu Leu Ile Gln Lys Gly Lys Asp Ile Lys Gly Val
           25           30           35
tcg gaa atc gtg cag aat ggg aag cac ttc aag ttc acc atc acc gct      260
Ser Glu Ile Val Gln Asn Gly Lys His Phe Lys Phe Thr Ile Thr Ala
           40           45           50
ggg tcc aaa gtg atc caa aac gaa ttc acg gtg ggg gag gaa tgt gag      308
Gly Ser Lys Val Ile Gln Asn Glu Phe Thr Val Gly Glu Glu Cys Glu
           55           60           65           70
ctg gag aca atg aca ggg gag aaa gtc aag aca gtg gtt cag ttg gaa      356
Leu Glu Thr Met Thr Gly Glu Lys Val Lys Thr Val Val Gln Leu Glu
           75           80           85
ggc gac aat aaa ctg gtg aca act ttc aaa aac atc aag tct gtg acc      404
Gly Asp Asn Lys Leu Val Thr Thr Phe Lys Asn Ile Lys Ser Val Thr
           90           95           100
gaa ctc aac ggc gac ata atc acc aat acc atg aca ttg ggt gac att      452
Glu Leu Asn Gly Asp Ile Ile Thr Asn Thr Met Thr Leu Gly Asp Ile
           105           110           115
gtc ttc aag aga atc agc aag aga att taaacaagtc tgcatttcat      499
Val Phe Lys Arg Ile Ser Lys Arg Ile
           120           125
attatttttag tgtgtaaaat taatgtaata aagtgaactt tggttttaa      547

```

<210> 422

<211> 669

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 29..259

<400> 422

```

gacgcakagc cgtggggaag ttttcgca atg gcc gcc gga acg tcg ccg gcc      52
                               Met Ala Ala Gly Thr Ser Pro Ala
                               1           5
gat gct ctc tgc gac gga aag aaa gcc gcc gga ccc act tcc ggt gcg      100
Asp Ala Leu Cys Asp Gly Lys Lys Ala Ala Gly Pro Thr Ser Gly Ala
           10           15           20
ttt tac aat gaa ctc cgg gca ctc gag tcg ctc ccc gga aac tat agt      148
Phe Tyr Asn Glu Leu Arg Ala Leu Glu Ser Leu Pro Gly Asn Tyr Ser
           25           30           35           40
tct gct tcc acc gtt gtc cta ctt tgg ata gca tca aga caa aaa tct      196
Ser Ala Ser Thr Val Val Leu Leu Trp Ile Ala Ser Arg Gln Lys Ser
           45           50           55
cag tct tcg agg aac tca cag tgt agc aga aaa gac aaa gga ggt aaa      244
Gln Ser Ser Arg Asn Ser Gln Cys Ser Arg Lys Asp Lys Gly Gly Lys
           60           65           70
cag gag gag cat gaa taattcatct tgatgtaatc atgcaagact cctaagaaga      299
Gln Glu Glu His Glu
           75
tgtaggaga ccctgtcatc attcttttgg cctctaattgt ccgttatctt catgtttgct      359

```

243

```

ggcttccttt tactactatg gaaatacaac atatgcaaaa gtagtggttt cgctgttttaa 419
aaacatcgtc agcagcatca tcatcattaa ctgcaccagg aagccatgac tcctttctgg 479
acaaggtgca cttgcctgtc atcccttccat ctctttcttc ctccttgaag ttcacaatac 539
ccagtccagg agcattgcag agaatgctag gtgtgtctgc agacttgact ttaaaagaaa 599
acaagttcct tcaagtgagt tcatcatgat ggggatgata tttgtttgag agaattgtgt 659
tctttggcaa 669

```

<210> 423
 <211> 546
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 78..302

```

<400> 423
ctygccagct gccgggtcttt cgggggctcc gtaactttct atccgtccgc gtcascyttg 60
ccaccctcat ctccaat atg cct ggt ccg acc ccc agt ggc act aac gtg 110
                Met Pro Gly Pro Thr Pro Ser Gly Thr Asn Val
                1             5             10
gga tcc tca ggg cgc tct ccc agc aaa gca gtg gcc gcc cgg gcg cgg 158
Gly Ser Ser Gly Arg Ser Pro Ser Lys Ala Val Ala Ala Arg Ala Arg
                15             20             25
gat cca ctg tcc ggc aga gga aaa atg cca gct gtg gga caa gga gtg 206
Asp Pro Leu Ser Gly Arg Gly Lys Met Pro Ala Val Gly Gln Gly Val
                30             35             40
cag gcc gca caa cct cgg cag gca ccg ggg gga tgt ggc gat tct aca 254
Gln Ala Ala Gln Pro Arg Gln Ala Pro Gly Gly Cys Gly Asp Ser Thr
                45             50             55
cag aag att cac ctg grc tca aag ttg gcc ctg ttc cag tat tgg tta 302
Gln Lys Ile His Leu Xaa Ser Lys Leu Ala Leu Phe Gln Tyr Trp Leu
                60             65             70             75
tgagtcttct gttcatcgct tctgtattta tgttgacat ttggggcaag tacactcggt 362
cgtagattca gttacatcca tctgtcatct gaagaakgag gaaawaaccc aacatttctt 422
ggacccaaaa gtatagtgcac tatctgttca tgagagaaat tttctgtaag cttgctgttt 482
ttaamcaggg gatttatcaa ttaattgatt tgaggaaatca gtttttttct atggctaata 542
aact 546

```

<210> 424
 <211> 499
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 16..183

```

<400> 424
tttattttgc cttct atg tca tct tgg agc tta atg tgc atg caa gat atg 51
                Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met
                1             5             10
gga aat act aaa gca aga cta ctc tat gaa gcc aat ctt cca gag aac 99
Gly Asn Thr Lys Ala Arg Leu Tyr Glu Ala Asn Leu Pro Glu Asn
                15             20             25
ttt cga aga cca cag aca gat caa gca gtg gaa ttt ttc atc aga gat 147
Phe Arg Arg Pro Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp
                30             35             40
aaa tat gaa aag aag aaa tac tac gaa aaa atg cca tagctattac 193
Lys Tyr Glu Lys Lys Lys Tyr Tyr Glu Lys Met Pro
                45             50             55
aatattttcc tcctctgatg ctcctcttca gcctttggta tcctctcctt ctctgcaagc 253
tgctgttgac aaaaataaat tggagaaaaga aaaggaaaaa aaaaaggaag agaaaaagag 313

```

244

```

agaaaaggag ccagaaaagc cggcaaaacc acttacagct gaaaagctgc agaagaaaga 373
tcagcaactg gagcctaaaa aaagtaccag ccctaaaaaa gctgcggasc cactgtggat 433
cttttaggac ttgtttggag aaagattgag aaggaaaaga gtgaaaggca gttagttaga 493
gtggaa 499

```

<210> 425
 <211> 512
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 122..487

```

<400> 425
aaacacatcc aagcttaaga cgggtgaggtc agcttcacat tctcaggaac tctccttctt 60
tgggtctggc tgaagttgag gatctcttac tctctaggcc acggaattaa cccgagcagg 120
c atg gag gcc tct gct ctc acc tca tca gca gtg acc agt gtg gcc aaa 169
Met Glu Ala Ser Ala Leu Thr Ser Ser Ala Val Thr Ser Val Ala Lys
1 5 10 15
gtg gtc agg gtg gcc tct ggc tct gcc gta gtt ttg ccc ctg gcc agg 217
Val Val Arg Val Ala Ser Gly Ser Ala Val Val Leu Pro Leu Ala Arg
20 25 30
att gct aca gtt gtg att gga gga gtt gtg gcc atg gcg gct gtg ccc 265
Ile Ala Thr Val Val Ile Gly Gly Val Val Ala Met Ala Ala Val Pro
35 40 45
atg gtg ctc agt gcc atg ggc ttc act gcg gcg gga atc gcc tcg tcc 313
Met Val Leu Ser Ala Met Gly Phe Thr Ala Ala Gly Ile Ala Ser Ser
50 55 60
tcc ata gca gcc aag atg atg tcc gcg gcg gcc att gcc aat ggg ggt 361
Ser Ile Ala Ala Lys Met Met Ser Ala Ala Ala Ile Ala Asn Gly Gly
65 70 75 80
gga gtt gcc tcg ggc agc ctt gtg gct act ctg cag tca ctg gga gca 409
Gly Val Ala Ser Gly Ser Leu Val Ala Thr Leu Gln Ser Leu Gly Ala
85 90 95
act gga ctc tcc gga ttg acc aag ttc atc ctg ggc tcc att ggg tct 457
Thr Gly Leu Ser Gly Leu Thr Lys Phe Ile Leu Gly Ser Ile Gly Ser
100 105 110
gcc att gcg gct gtc att gcg agg ttc tac tagctccctg ccctcgctg 507
Ala Ile Ala Ala Val Ile Ala Arg Phe Tyr
115 120
cagag 512

```

<210> 426
 <211> 502
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 57..278

```

<400> 426
attcaccggc gacgcgatac gggtcctcca ccgaggccca tgcgaagctt tccact atg 59
Met
1
gct tcc agc act gtc ccg gtg agc gct gct ggc tcg gct aat gaa act 107
Ala Ser Ser Thr Val Pro Val Ser Ala Ala Gly Ser Ala Asn Glu Thr
5 10 15
ccc gaa ata ccg gac aac gtg gga gat tgg ctt cgg ggc gtc tac cgc 155
Pro Glu Ile Pro Asp Asn Val Gly Asp Trp Leu Arg Gly Val Tyr Arg
20 25 30
ttt gcc act gat agg aat gac ttc cgg agg aac ttg ata cta aat ttg 203

```

245

Phe Ala Thr Asp Arg Asn Asp Phe Arg Arg Asn Leu Ile Leu Asn Leu
 35 40 45
 gga ctc ttt gct gcg gga gtt tgg ctg gcc agg aac ttg agt gac att 251
 Gly Leu Phe Ala Ala Gly Val Trp Leu Ala Arg Asn Leu Ser Asp Ile
 50 55 60 65
 gac ctc atg gca cct cag cca ggg gtg tagccaagta gttctaattgc 298
 Asp Leu Met Ala Pro Gln Pro Gly Val
 70
 cacctgtcgt cttatcatct gattgcagac aaatggaatc ctgtgctgaa cccgaatcctt 358
 ccaaaaaaca gcctacaatc tgtgaccacc acaagatgtg ccctgatggc agctgaagtt 418
 tgattcagat gggcactttt cttccccttc cctgcctagt ttccttttgt tccttgagtc 478
 acgcagaatt ccattctctg gtca 502

<210> 427

<211> 575

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 50..340

<400> 427

gaggagaaga ggtgaccccg ctaaagataa taaatcatca ttgatcaag atg tct tca 58
 Met Ser Ser
 1
 gca cct gag cct cca aca ttc aaa aag gaa cca ccc aaa gar aaa gag 106
 Ala Pro Glu Pro Pro Thr Phe Lys Lys Glu Pro Pro Lys Glu Lys Glu
 5 10 15
 ttt caa agc cca ggg ctc aga ggg gtg cgc aca aca acc tta ttt cgt 154
 Phe Gln Ser Pro Gly Leu Arg Gly Val Arg Thr Thr Thr Leu Phe Arg
 20 25 30 35
 gct gtg aat cca gag ctc ttc att aaa cct aac aaa cct gta atg gct 202
 Ala Val Asn Pro Glu Leu Phe Ile Lys Pro Asn Lys Pro Val Met Ala
 40 45 50
 ttc gga ttg gta act ctt tca ctt tgc gtg gca tat att ggt tat cta 250
 Phe Gly Leu Val Thr Leu Ser Leu Cys Val Ala Tyr Ile Gly Tyr Leu
 55 60 65
 cat gca ata caa gag aat aaa aag gac ctc tat gaa gct att gat agt 298
 His Ala Ile Gln Glu Asn Lys Lys Asp Leu Tyr Glu Ala Ile Asp Ser
 70 75 80
 gag ggg cac agt tat atg agg cgg aaa aca tcc aaa tgg gat 340
 Glu Gly His Ser Tyr Met Arg Arg Lys Thr Ser Lys Trp Asp
 85 90 95
 tagtagtgct ggtagtgca ratggacctt tattaaggt tctgaaatct tcaaatgaaa 400
 gaccttgatg gtgtacagta tcatgtttct tggtctagaa catgctaattg aagagagaag 460
 atagcagttg caaccagaca actgtcgtaa attttgtcct ttcacagctg cagccattat 520
 ctcattcttt ttccacagag tgagcgtcat aatattttct ttccttactc ttata 575

<210> 428

<211> 493

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 177..470

<400> 428

atcgaccacg gaaatttgac acctccgggc ttggaagcag ctctctcctc cttccccgct 60
 gcttataaac ctcagccctg aggcctcagc tcaactctacc ccatctcctt gccgggtcag 120
 ccctgacaaa ggtcagctag ccccttgagg acatcagctt tggcctcagg gtccta atg 179
 Met

1
gca gca gaa cca ctg aca gag cta gag gag tcc att gag aac gtg gtc 227
Ala Ala Glu Pro Leu Thr Glu Leu Glu Glu Ser Ile Glu Asn Val Val
5 10 15
acc acc ttc ttc acc ttt gca agg cag gag ggc cgg aag gat agc ctc 275
Thr Thr Phe Phe Thr Phe Ala Arg Gln Glu Gly Arg Lys Asp Ser Leu
20 25 30
agc gtc aac gag ttc aaa gag ctg gtt acc cag cag ttg ccc cat ctg 323
Ser Val Asn Glu Phe Lys Glu Leu Val Thr Gln Gln Leu Pro His Leu
35 40 45
ctc aag gat gtg ggc tct ctt gat gag aag atg aag agc ttg gat gtg 371
Leu Lys Asp Val Gly Ser Leu Asp Glu Lys Met Lys Ser Leu Asp Val
50 55 60 65
aat cag gac tcg gag ctc aag ttc aat gag tac tgg aga ttg att ggg 419
Asn Gln Asp Ser Glu Leu Lys Phe Asn Glu Tyr Trp Arg Leu Ile Gly
70 75 80
gag ctg gcc aag gaa atc agg aag aag aaa gac ctk aag atc agg aag 467
Glu Leu Ala Lys Glu Ile Arg Lys Lys Lys Asp Leu Lys Ile Arg Lys
85 90 95
aag taaagccgcc tggctgagat ggg 493
Lys

<210> 429
<211> 505
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 114..458

<400> 429
agttccggct ctgcstgaag agctttgcat tgtgggaagt ytttytttct cgttccccgg 60
ccatcttagc ggctgctggt ggttgggggc cgtcccgcctc ctaaggcagg aag atg 116
Met
1
gtg gcc gca aag aag acg aaa aag tcg ctg gag tcg atc aac tct agg 164
Val Ala Ala Lys Lys Thr Lys Lys Ser Leu Glu Ser Ile Asn Ser Arg
5 10 15
ctc caa ctc gtt atg aaa agt ggg aag tac gtc ctg ggg tac aag cag 212
Leu Gln Leu Val Met Lys Ser Gly Lys Tyr Val Leu Gly Tyr Lys Gln
20 25 30
act ctg aag atg atc aga caa ggc aaa gcg aaa ttg gtc att ctc gct 260
Thr Leu Lys Met Ile Arg Gln Gly Lys Ala Lys Leu Val Ile Leu Ala
35 40 45
aac aac tgc cca gct ttg agg aaa tct gaa ata gag tac tat gct atg 308
Asn Asn Cys Pro Ala Leu Arg Lys Ser Glu Ile Glu Tyr Tyr Ala Met
50 55 60 65
ttg gct aaa act ggt gtc cat cac tac agt ggc aat aat att gaa ctg 356
Leu Ala Lys Thr Gly Val His His Tyr Ser Gly Asn Asn Ile Glu Leu
70 75 80
ggc aca gca tgc gga aaa tac tac aga gtg tgc aca ctg gct atc att 404
Gly Thr Ala Cys Gly Lys Tyr Tyr Arg Val Cys Thr Leu Ala Ile Ile
85 90 95
gat cca ggt gac tct gac atc att aga agc atg cca gaa cag act ggt 452
Asp Pro Gly Asp Ser Asp Ile Ile Arg Ser Met Pro Glu Gln Thr Gly
100 105 110
gaa aag taaacctttt cacctacaaa atttcacctg caaaaaaaaaa aaaaaav 505
Glu Lys
115

<210> 430
<211> 772

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 324..629

<400> 430
ataatcgcgt catttcgagg aggggacgaa ggggtagttc tttcacctcg gctgggagcc 60
tagaaaagcc tagaaacagc tccttttttc ttccgcctcc gagtcttcgc gtcagcgccc 120
tgcgcagggc ccttggggcg aatcgcggtg cgcgtcgggg cgaccgccct ccctccctgg 180
gaggggagag ggggctagcg gcgacsgctg gggcgagcgc gcctgcgcgc tgggtgattt 240
tttcacgtgt cgccaggggc ggactgcgag tctcttttgc gcgctacact agagcagagt 300
acgagtctga ggcggaggga gta atg gca gga caa gcg ttt aga aag ttt ctt 353
Met Ala Gly Gln Ala Phe Arg Lys Phe Leu
1 5 10
cca ctc ttt gac cga gta ttg gtt gaa agg agt gct gct gaa act gta 401
Pro Leu Phe Asp Arg Val Leu Val Glu Arg Ser Ala Ala Glu Thr Val
15 20 25
acc aaa gga ggc att atg ctt cca gaa aaa tct caa gga aaa gta ttg 449
Thr Lys Gly Gly Ile Met Leu Pro Glu Lys Ser Gln Gly Lys Val Leu
30 35 40
caa gca aca gta gtc gct gtt gga tcg ggt tct aaa gga aag ggt gga 497
Gln Ala Thr Val Val Ala Val Gly Ser Gly Ser Lys Gly Lys Gly Gly
45 50 55
gag att caa cca gtt agc gtg aaa gtt gga gat aaa gtt ctt ctc cca 545
Glu Ile Gln Pro Val Ser Val Lys Val Gly Asp Lys Val Leu Leu Pro
60 65 70
gaa tat gga ggc acc aaa gta gtt cta gat rac aag gat tat ttc cta 593
Glu Tyr Gly Gly Thr Lys Val Val Leu Asp Xaa Lys Asp Tyr Phe Leu
75 80 85 90
ttt aga gat ggt gac att ctt gga aag tac gta gac tgaaataagt 639
Phe Arg Asp Gly Asp Ile Leu Gly Lys Tyr Val Asp
95 100
cactattgaa atggcatcaa catgatgctg cccattccac tgaagttctg aaatctttcg 699
tcatgtaaat aatttccata tttctctttt ataataaact aatgataact aatgaccaa 759
aaaaaaaaa aaa 772

<210> 431
<211> 672
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 224..529

<400> 431
cataagaaca gggagggttag aagtagggctc ttggtgacaa aatatgttgt gtagagttca 60
ggggagagtg cgtcatatgt tgttctagga agattgtagt ggtgaggggtg tttattataa 120
taatgtttgt gtattcggct atgaagaata ggccaggggc ggactgcgag tctcttttgcg 180
gcgctacact agagcagagt acgagtctga ggcggaggga gta atg gca gga caa 235
Met Ala Gly Gln
1
gcg ttt aga aag ttt ctt cca ctc ttt gac cga gta ttg gtt gaa agg 283
Ala Phe Arg Lys Phe Leu Pro Leu Phe Asp Arg Val Leu Val Glu Arg
5 10 15 20
agt gct gct gaa act gta acc aaa gga ggc att atg ctt cca gaa aaa 331
Ser Ala Ala Glu Thr Val Thr Lys Gly Gly Ile Met Leu Pro Glu Lys
25 30 35
tct caa gga aaa gta ttg caa gca aca gta gtc gct gtt gga tcg ggt 379
Ser Gln Gly Lys Val Leu Gln Ala Thr Val Val Ala Val Gly Ser Gly
40 45 50

248

tct	aaa	gga	aag	ggt	gga	gag	att	caa	cca	ggt	agc	gtg	aaa	ggt	gga	427
Ser	Lys	Gly	Lys	Gly	Gly	Glu	Ile	Gln	Pro	Val	Ser	Val	Lys	Val	Gly	
	55					60					65					
gat	aaa	ggt	ctt	ctc	cca	gaa	tat	gga	ggc	acc	aaa	gta	ggt	cta	gat	475
Asp	Lys	Val	Leu	Leu	Pro	Glu	Tyr	Gly	Gly	Thr	Lys	Val	Val	Leu	Asp	
	70					75				80						
gac	aag	gat	tat	ttc	cta	ttt	aga	gat	ggt	gac	att	ctt	ggr	aag	tac	523
Asp	Lys	Asp	Tyr	Phe	Leu	Phe	Arg	Asp	Gly	Asp	Ile	Leu	Gly	Lys	Tyr	
	85			90					95			100				
gta	gac	tga	aata	aagt	cact	att	tgaa	atgg	cat	caa	cat	gat	gct	g	ccatt	ccac
Val	Asp															579
tga	agtt	ctg	aaat	cttt	cg	tcat	gtaa	aat	ttt	ccata	ttt	ctct	ttt	ata	ataa	aact
aat	gata	aact	aat	gac	caaa	aaaa	aaaaaa	aaa								639
																672

<210> 432
 <211> 517
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 97..444

at	cttt	ccgt	ccc	ggg	cagc	cagc	gcc	agc	cg	gag	cc	cagc	gcc	gcc	ccat	ca	60
ct	gcc	gct	gc	caag	tcc	acc	gct	gcc	ccc	gcc	atg	tct	gct	acc	gct	gcc	114
										Met	Ser	Ala	Thr	Ala	Ala		
										1					5		
ac	g	cc	ccc	cct	gct	gcc	ccg	gct	ggg	gag	ggt	ggt	ccc	cct	gca	ccc	162
Thr	Ala	Pro	Pro	Ala	Ala	Pro	Ala	Gly	Glu	Gly	Gly	Pro	Pro	Ala	Pro		
			10					15					20				
cct	cca	aac	ctc	acc	agt	aac	agg	aga	ctg	cag	cag	acc	cag	gcc	cag	210	
Pro	Pro	Asn	Leu	Thr	Ser	Asn	Arg	Arg	Leu	Gln	Gln	Thr	Gln	Ala	Gln		
		25				30					35						
gtg	gat	gag	gtg	gtg	gac	atc	atg	agg	gtg	aac	gtg	gac	aag	gtc	ctg	258	
Val	Asp	Glu	Val	Val	Asp	Ile	Met	Arg	Val	Asn	Val	Asp	Lys	Val	Leu		
	40				45					50							
gag	cga	gac	cag	aag	ctg	tcg	gag	ctg	gac	gac	cgt	gca	gat	gca	ctc	306	
Glu	Arg	Asp	Gln	Lys	Leu	Ser	Glu	Leu	Asp	Asp	Arg	Ala	Asp	Ala	Leu		
	55			60					65						70		
cag	gcg	ggg	gcc	tcc	cag	ttt	gaa	aca	agc	gca	gcc	aag	ctc	aag	cg	354	
Gln	Ala	Gly	Ala	Ser	Gln	Phe	Glu	Thr	Ser	Ala	Ala	Lys	Leu	Lys	Arg		
			75					80					85				
aaa	tac	tgg	tgg	aaa	aac	ctc	aag	atg	atc	atc	atc	ttg	gga	gtg	att	402	
Lys	Tyr	Trp	Trp	Lys	Asn	Leu	Lys	Met	Ile	Ile	Ile	Leu	Gly	Val	Ile		
		90					95					100					
tgc	gcc	atc	atc	ctc	atc	atc	atc	ata	gtt	tac	ttc	agc	act			444	
Cys	Ala	Ile	Ile	Leu	Ile	Ile	Ile	Ile	Val	Tyr	Phe	Ser	Thr				
	105				110						115						
t	aaat	cccc	g	agg	agt	ctg	c	ctg	c	ctaga	ga	aggg	c	ctc	tcccc	caacc	
c	ctcc	ac	ctc	tct												504	
																517	

<210> 433
 <211> 485
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 21..311

ttt	gtt	ttt	ct	aat	agg	gcaa	atg	tct	cta	agc	ttg	gtg	ttt	aga	gct	gct	tca	53
-----	-----	-----	----	-----	-----	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

249

	Met	Ser	Leu	Ser	Leu	Val	Phe	Arg	Ala	Ala	Ser	
	1				5				10			
tat ttt aaa cta gtt cca ttc cac agt tct agt tca aac cag ttt tta												101
Tyr Phe Lys Leu Val Pro Phe His Ser Ser Ser Ser Asn Gln Phe Leu												
	15				20				25			
cag cct cct ggg tgg gtc gtc ttg acc caa act ctt gtg ttg tta cat												149
Gln Pro Pro Gly Trp Val Val Leu Thr Gln Thr Leu Val Leu Leu His												
	30				35				40			
ttt gag agg ttt tca tac cag aat gta cca raa agt gca caa ggt aaa												197
Phe Gln Arg Phe Ser Tyr Gln Asn Val Pro Xaa Ser Ala Gln Gly Lys												
	45				50				55			
ggt aat tta cag cca gaa aca aat ata cat ttg ttt cat ttc ttg act												245
Gly Asn Leu Gln Pro Glu Thr Asn Ile His Leu Phe His Phe Leu Thr												
	60				65				70			75
ttc cct aag cag ata agc aga aac ttg ttt aac tca tta ctt tgt ttg												293
Phe Pro Lys Gln Ile Ser Arg Asn Leu Phe Asn Ser Leu Leu Cys Leu												
	80				85				90			
atg tgt ctt aca tat ttt tgactaaaag ttatagaatg ttattcctct												341
Met Cys Leu Thr Tyr Phe												
	95											
gggggaatct tttacaggtg gaggaatggg gatagcagta ttgcctcaga ttcaaactgg												401
catcaccata aaccctgtag gccagggtgg aatgaagtca gctccttttt atagttgaaa												461
tacaattttt tattcaccat tgtc												485

<210> 434

<211> 511

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 31..369

<400> 434

acttgctctg cccgctcaaa cagttgcagg atg tcg atg aca gac ttg ctg aac												54
	Met	Ser	Met	Thr	Asp	Leu	Leu	Asn				
	1								5			
gct gag gac atc aag aag gcg gtg gga gcc ttt agc gct acc gac tcc												102
Ala Glu Asp Ile Lys Lys Ala Val Gly Ala Phe Ser Ala Thr Asp Ser												
	10				15				20			
ttc gac cac aaa aag ttc ttc caa atg gtc ggc ctg aag aaa aag agt												150
Phe Asp His Lys Lys Phe Phe Gln Met Val Gly Leu Lys Lys Lys Ser												
	25				30				35			40
gcg gat gat gtg aag aag gtg ttt cac atg ctg gac aag gac aaa agt												198
Ala Asp Asp Val Lys Lys Val Phe His Met Leu Asp Lys Asp Lys Ser												
	45				50				55			
ggc ttc atc gag gag gat gag ctg gga ttc atc cta aaa ggc ttc tcc												246
Gly Phe Ile Glu Glu Asp Glu Leu Gly Phe Ile Leu Lys Gly Phe Ser												
	60				65				70			
cca gat gcc aga gac ctg tct gct aaa gaa acc aag atg ctg atg gct												294
Pro Asp Ala Arg Asp Leu Ser Ala Lys Glu Thr Lys Met Leu Met Ala												
	75				80				85			
gct gga gac aaa gat ggg gac ggc aaa att ggg gtt gac gaa ttc tcc												342
Ala Gly Asp Lys Asp Gly Asp Gly Lys Ile Gly Val Asp Glu Phe Ser												
	90				95				100			
act ctg gtg gct gaa ast aag aag cac tgactgcccc tgggtcttcca												389
Thr Leu Val Ala Glu Xaa Lys Lys His												
	105				110							
cctctctgcc ctgaacaccc aatctcggcc cctctcgcca cctcctgca tttctgttca												449
gttcgtttat gttatttttt actcccccct cccctgtggc cctctaataga caccattctt												509
ct												511

<210> 435

```
<211> 511
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 82..417
```

[illegible]

```
<210> 436
<211> 475
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 22..375
```

<400> 436																
agtcgagcca	g	atg	gtg	ttg	ttg	gag	agc	gag	cag	ttc	ctg					51
		Met	Val	Leu	Leu	Glu	Ser	Glu	Gln	Phe	Leu					
		1				5					10					
acg	gag	ctg	acc	aga	ctt	ttc	cag	aag	tgc	cgg	acg	tcg	ggc	agc	gtc	99
Thr	Glu	Leu	Thr	Arg	Leu	Phe	Gln	Lys	Cys	Arg	Thr	Ser	Gly	Ser	Val	
			15						20					25		
tat	atc	acc	ttg	aag	aag	tat	gac	ggt	cga	acc	aaa	ccc	att	cca	aag	147
Tyr	Ile	Thr	Leu	Lys	Lys	Tyr	Asp	Gly	Arg	Thr	Lys	Pro	Ile	Pro	Lys	
			30					35					40			
aaa	ggt	act	gtg	gag	ggc	ttt	gag	ccc	gca	gac	aac	aag	tgt	ctg	tta	195
Lys	Gly	Thr	Val	Glu	Gly	Phe	Glu	Pro	Ala	Asp	Asn	Lys	Cys	Leu	Leu	
			45				50					55				
aga	gct	acc	gat	ggg	aag	aag	aag	atc	agc	act	gtg	gtg	agc	tcc	aag	243
Arg	Ala	Thr	Asp	Gly	Lys	Lys	Lys	Ile	Ser	Thr	Val	Val	Ser	Ser	Lys	
	60					65					70					
gaa	gtg	aat	aag	ttt	cag	atg	gct	tat	tca	aac	ctc	ctt	aga	gct	aac	291

251

Glu Val Asn Lys Phe Gln Met Ala Tyr Ser Asn Leu Leu Arg Ala Asn
 75 80 85 90
 atg gat ggg ttg aag aag aga gac aaa aag aac gaa act aag aag acc 339
 Met Asp Gly Leu Lys Lys Arg Asp Lys Lys Asn Glu Thr Lys Lys Thr
 95 100 105
 aaa gca gca gca gca gca gca gca aca gca gca cag taaagggcat 385
 Lys Ala Ala Ala Ala Ala Ala Thr Ala Ala Gln
 110 115
 acatttcctg ctttcaccaa ttaaccactg aattgctatt ttttcctttt ggccagatag 445
 ctaggtttct ggttccccca cagtaggtgt 475

<210> 437
 <211> 896
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 351..677

<220>
 <221> misc_feature
 <222> 234
 <223> n=a, g, c or t

<400> 437
 gcacgagacc attacataaa aacaaacagg aatgcattct tgacatttcc gaacacacat 60
 taagtgaaaa tgacttagaa gaactaagg tagatcacta taaatgtaac atacaggcat 120
 ctgtacatgt ttctgatttc agtacagata atagtggatc tcaaccaaaa cagaagtcag 180
 atactgtgct ttttcagca aaggatctca aggaaaagga ctttcattca atanttactc 240
 atgattctgg tctgataaca ataaacagtt cacaagagca cctaactgtt caggcaaagg 300
 ctccattcca tactcctcct gaggaaccca atgaatgtga cttcaagaat atg gat 356
 Met Asp
 1
 agt tta cct tct ggt aaa ata crt cga aaa gtg aaa ata ata tta gga 404
 Ser Leu Pro Ser Gly Lys Ile Xaa Arg Lys Val Lys Ile Ile Leu Gly
 5 10 15
 cga aat aga aaa gaa aat ctg gaa cca aat gct gaa ttt gat aaa aga 452
 Arg Asn Arg Lys Glu Asn Leu Glu Pro Asn Ala Glu Phe Asp Lys Arg
 20 25 30
 act gaa ttt wtt aca cra gaa gaa aac aga att tgt agt tca ccg gta 500
 Thr Glu Phe Xaa Thr Xaa Glu Glu Asn Arg Ile Cys Ser Ser Pro Val
 35 40 45 50
 cag tct tta cta gac ttg ttt cag act agt gaa gag aaa tca gaa ttt 548
 Gln Ser Leu Leu Asp Leu Phe Gln Thr Ser Glu Glu Lys Ser Glu Phe
 55 60 65
 ttg ggt ttc aca agc tac aca gaa aag agt ggt ata tgc aat gtt tta 596
 Leu Gly Phe Thr Ser Tyr Thr Glu Lys Ser Gly Ile Cys Asn Val Leu
 70 75 80
 gat att tgg gaa gag gaa aat tca gat aat ctg tta aca gcg ttw ttc 644
 Asp Ile Trp Glu Glu Glu Asn Ser Asp Asn Leu Leu Thr Ala Xaa Phe
 85 90 95
 tcg tcc cct tca act tct aca ttt act ggc ttt tagaatttaa aaaatgcata 697
 Ser Ser Pro Ser Thr Ser Phe Thr Gly Phe
 100 105
 cttttcagaa gtgataagga tcatattctt gaaattttta taaatatgta tggaaattct 757
 taggattttt ttaccagctt tgtttacaga cccaaatgta aatattaaaa ataaatattt 817
 gcaattttct acagaattga atacctgtta aagaaaaatt acagaataaa cttgtgactg 877
 gtaaaaaaaaa aaaaaaatg 896

<210> 438
 <211> 576
 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 162..470

<400> 438

```

cattttaatt catatytgag tgggcggtgg cgattggtgt tggcggtctg gctcagctgg      60
gcagggggta actttactga tttgggggtg gtttttagtt taatttttct tttctagctt      120
cccacgcacg gtcagtgcgc acgttgtaat cagctgaggc c atg tca gga gac gga      176
                                   Met Ser Gly Asp Gly
                                   1           5

gcc acg gag cag gca gct gag tat gtc cca gag aag gtg aag aaa gcg      224
Ala Thr Glu Gln Ala Ala Glu Tyr Val Pro Glu Lys Val Lys Lys Ala
                                   10           15           20

gaa aag aaa tta gaa gag aat cca tat gac ctt gat gct tgg agc att      272
Glu Lys Lys Leu Glu Glu Asn Pro Tyr Asp Leu Asp Ala Trp Ser Ile
                                   25           30           35

ctc att cga gag gca cag aat caa cct ata gac aaa gca cgg aag act      320
Leu Ile Arg Glu Ala Gln Asn Gln Pro Ile Asp Lys Ala Arg Lys Thr
                                   40           45           50

tat gaa cgc ctt gtt gcc cag ttc ccc agt tct ggc aga ttc tgg aaa      368
Tyr Glu Arg Leu Val Ala Gln Phe Pro Ser Ser Gly Arg Phe Trp Lys
                                   55           60           65

ctg tac att gaa gca gag gtt act att tta ttt tat ttt ttc tta tat      416
Leu Tyr Ile Glu Ala Glu Val Thr Ile Leu Phe Tyr Phe Phe Leu Tyr
                                   70           75           80           85

cag tat tgc agc att cac tgt agt gat aga aaa caa gtt agg aac ata      464
Gln Tyr Cys Ser Ile His Cys Ser Asp Arg Lys Gln Val Arg Asn Ile
                                   90           95           100

gcc aat taggacaagg aggatttaaa tgtgtcttac ctttattttg taaaataggt      520
Ala Asn
ataaaggagt aattaaaatg aatttttgaa atttgggtct tttacaagct gatgat      576

```

<210> 439

<211> 484

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 32..349

<400> 439

```

actaactagc aaacggggac tagaaatagg g atg ctg aaa agc aac ggg gag      52
                                   Met Leu Lys Ser Asn Gly Glu
                                   1           5

aga cgc agt cgt aac gca ctt ccg gcg gtc tac gcg agg aag atg gct      100
Arg Arg Ser Arg Asn Ala Leu Pro Ala Val Tyr Ala Arg Lys Met Ala
                                   10           15           20

gca tcc cag cag caa gct tca gcg gct tcc tca gct gct ggt gta tcg      148
Ala Ser Gln Gln Gln Ala Ser Ala Ala Ser Ser Ala Ala Gly Val Ser
                                   25           30           35

ggt cct agt tcg gct ggc ggc ccg ggt ccc cag cag cag ccg caa ccg      196
Gly Pro Ser Ser Ala Gly Gly Pro Gly Pro Gln Gln Gln Pro Gln Pro
                                   40           45           50           55

cca gca caa ctg gtg ggc cct gcc cag agc ggc ctc ctg cag caa cag      244
Pro Ala Gln Leu Val Gly Pro Ala Gln Ser Gly Leu Leu Gln Gln Gln
                                   60           65           70

caa cag gac ttc gat cct gtg cag cgt tat aag atg ctc atc ccg cag      292
Gln Gln Asp Phe Asp Pro Val Gln Arg Tyr Lys Met Leu Ile Pro Gln
                                   75           80           85

ctg aag gag agt cta cag gtg att ggc ctt aag cag cga gaa gca aac      340

```

253

Leu Lys Glu Ser Leu Gln Val Ile Gly Leu Lys Gln Arg Glu Ala Asn
 90 95 100
 tgg att tgg tagtctagag gggcaggccg rgggccaggt tagtcggaga 389
 Trp Ile Trp
 105
 gagcgctaag cagaaagagg aatagaacct gctgttttgg cctggctggt gcacgcctgt 449
 gttccagcac ttggggaggc cgaggcgkgc ggatc 484

<210> 440
 <211> 523
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 114..371

<400> 440
 ctctctcagc ttccggctgg tagtagttcc gcttcctgtc cgactgtggt gtctttgctg 60
 agggtcacat tgagctgcag gttgaatccg ggggtgccttt aggattcagc acc atg 116
 Met
 1
 gcg gaa gac atg gag acc aaa atc aag aac tac aag acc gcc cct ttt 164
 Ala Glu Asp Met Glu Thr Lys Ile Lys Asn Tyr Lys Thr Ala Pro Phe
 5 10 15
 gac agc cgc ttc ccc aac cag aac cag act aga aac tgc tgg cag aac 212
 Asp Ser Arg Phe Pro Asn Gln Asn Gln Thr Arg Asn Cys Trp Gln Asn
 20 25 30
 tac ctg gac ttc cac cgc tgt cag aag gca atg acc gct aaa gga ggc 260
 Tyr Leu Asp Phe His Arg Cys Gln Lys Ala Met Thr Ala Lys Gly Gly
 35 40 45
 gat atc tct gtg tgc gaa tgg tac cag cgt gtg tac cag tcc ctc tgc 308
 Asp Ile Ser Val Cys Glu Trp Tyr Gln Arg Val Tyr Gln Ser Leu Cys
 50 55 60 65
 ccc aca tcc tgg gtc aca gac tgg gat gag caa cgg gct gaa ggc acg 356
 Pro Thr Ser Trp Val Thr Asp Trp Asp Glu Gln Arg Ala Glu Gly Thr
 70 75 80
 ttt ccc ggg aag atc tgaactggct gcattctcct ttccctctgtc ctccatcctt 411
 Phe Pro Gly Lys Ile
 85
 ctcccaggat ggtgaagggg gacctgttac ccagtgatcc ccaccccagg atcctaaatc 471
 atgacttacc tgctaataaaa aactcattgg aaaagyraaa aaaaaaaaaa aa 523

<210> 441
 <211> 499
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 95..409

<400> 441
 ttccaagtgg ttgtstctgc actatttact gaaaagttca tcctttcccc agtgatttgt 60
 aacactgcct ctttcataaa ttaaattttt gtgt atg tgt gtg ggt ctt ttg atg 115
 Met Cys Val Gly Leu Leu Met
 1 5
 gtt tct att ctg act gac atc aat ttg tct aat ctt gta gca gta cag 163
 Val Ser Ile Leu Thr Asp Ile Asn Leu Ser Asn Leu Val Ala Val Gln
 10 15 20
 tac agt cct gat tat tgc aac ttt agg aaa agg tct gat aaa aac caa 211
 Tyr Ser Pro Asp Tyr Cys Asn Phe Arg Lys Arg Ser Asp Lys Asn Gln
 25 30 35

254

```

gat gcc tcc aca ttt tgt cat aat tgt aac caa ttt cac ctc tgt ctc      259
Asp Ala Ser Thr Phe Cys His Asn Cys Asn Gln Phe His Leu Cys Leu
40                               45                               50                               55
cag tat cac cac aaa att gtt ctt cct tgg agt gtc ttg gct att ctt      307
Gln Tyr His His Lys Ile Val Leu Pro Trp Ser Val Leu Ala Ile Leu
                               60                               65                               70
agc caa ctg ttc ctc cat att act tct aga att aga cca aca att tat      355
Ser Gln Leu Phe Leu His Ile Thr Ser Arg Ile Arg Pro Thr Ile Tyr
                               75                               80                               85
aaa tca aac aaa ccc aag agc att gaa att ttg att gga ttt gta ttg      403
Lys Ser Asn Lys Pro Lys Ser Ile Glu Ile Leu Ile Gly Phe Val Leu
                               90                               95                               100
aat tta tagattaatc tggtaaacca tgtcatcttt acaatgttgt cttccaatac      459
Asn Leu
105
atgaatatgg tacagctctt cattttactta ggccctttgaa      499

```

<210> 442

<211> 460

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 75..377

<400> 442

```

ccctcgcgga gcttactgag cgcggccccc agcccagctc cgccgcccag cgccctgtgcc      60
ggcacggmwa cacc atg gag cgc ccg gat aag gcg gcg ctg aac gca ctg      110
                               Met Glu Arg Pro Asp Lys Ala Ala Leu Asn Ala Leu
                               1                               5                               10
cag cct cct gag ttc aga aat gaa agc tca tta gca tct aca ctg aag      158
Gln Pro Pro Glu Phe Arg Asn Glu Ser Ser Leu Ala Ser Thr Leu Lys
                               15                               20                               25
acg ctc ctg ttc ttc aca gct tta atg atc act gtt cct att ggg tta      206
Thr Leu Leu Phe Phe Thr Ala Leu Met Ile Thr Val Pro Ile Gly Leu
                               30                               35                               40
tat ttc aca act aaa tct tac ata ttt gaa ggc gcc ctt ggg atg tcc      254
Tyr Phe Thr Thr Lys Ser Tyr Ile Phe Glu Gly Ala Leu Gly Met Ser
45                               50                               55                               60
aat agg gac agc tat ttt tac gct gct att gtt gca gtg gtc gcc gtc      302
Asn Arg Asp Ser Tyr Phe Tyr Ala Ala Ile Val Ala Val Val Ala Val
                               65                               70                               75
cat gtg gtg ctg gcc ctc ttt gtg tat ktg gcc tgg aat gaa ggc tca      350
His Val Val Leu Ala Leu Phe Val Tyr Xaa Ala Trp Asn Glu Gly Ser
                               80                               85                               90
cga cag tgc cgt gaa ggc aaa cag gat taaagtgaac atcacctttt      397
Arg Gln Xaa Arg Glu Gly Lys Gln Asp
95                               100
tatagcatta aattcatttt ttaaaatgat aaatgctgga gggggccatc tgatttgaat      457
aaa      460

```

<210> 443

<211> 470

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 30..305

<400> 443

```

agaatgtcgt actctcagga ggctccacc atg ttc agg gat ttc gga cgc cga      53

```

255

	Met	Phe	Arg	Asp	Phe	Gly	Arg	Arg	
	1				5				
ctg cag agg gat ttg aag aga gtg gtg gat gct agg ctg agg ctc agc									101
Leu Gln Arg Asp Leu Lys Arg Val Val Asp Ala Arg Leu Arg Leu Ser									
10 15 20									
gag gag ctc agc ggc ggg agg atc aag ccg aag cct gtg gag gtc cag									149
Glu Glu Leu Ser Gly Gly Arg Ile Lys Pro Lys Pro Val Glu Val Gln									
25 30 35 40									
gtg gtc acg cat cac atg cag cgc tac gcc gtg tgg ttc gga ggc tcc									197
Val Val Thr His His Met Gln Arg Tyr Ala Val Trp Phe Gly Gly Ser									
45 50 55									
atg ctg gcc tcg act ccc gag ttc ttt cag gtc tgc cac acc aag aag									245
Met Leu Ala Ser Thr Pro Glu Phe Phe Gln Val Cys His Thr Lys Lys									
60 65 70									
gac tat gaa gag tac ggg ccc agc atc tgc cgc cac aac ccc gtc ttt									293
Asp Tyr Glu Glu Tyr Gly Pro Ser Ile Cys Arg His Asn Pro Val Phe									
75 80 85									
gga gtc atg tcc tagtgtctgc ctgaacgcgt cgttcgatgg tgtcacgttg									345
Gly Val Met Ser									
90									
gggaacaagt gtccttcaga acccagagaa ggccgccgtt ctgtaaatag cgacgtcggg									405
ggtgctgccc agcagcgtgc ttgcattgcc ggtgcatgag gcgcggcgcg ggccttcagt									465
aaaag									470

<210> 444

<211> 521

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 162..455

<400> 444

agaatacgtgta caccaatgcg gacaagctcc tagaagcagc agagcagttg gctcagacgg									60
gggaatgtga ccccgaggag atctacaagg cagctcgaca cctggagggtg cgcattccaag									120
acttcgtgcg cagggtggag cagcggaagc ttctcctgga c atg tct gtt tcc ttc									176
					Met	Ser	Val	Ser	Phe
					1				5
cac aca cac acc aaa gag ttg tgg aca tgg atg gaa gac ctt cag aag									224
His Thr His Thr Lys Glu Leu Trp Thr Trp Met Glu Asp Leu Gln Lys									
10 15 20									
gag atg ttg gag gat gtc tgt gca gat tct gtg gat gca gtc cag gaa									272
Glu Met Leu Glu Asp Val Cys Ala Asp Ser Val Asp Ala Val Gln Glu									
25 30 35									
ctg atc aag cag ttc cag cag cag cag acc gcc act cta kat gcc aca									320
Leu Ile Lys Gln Phe Gln Gln Gln Gln Thr Ala Thr Leu Xaa Ala Thr									
40 45 50									
ctc aat gtc atc aag gaa ggc gaa gac ctt atc cag cag ctc agg tca									368
Leu Asn Val Ile Lys Glu Gly Glu Asp Leu Ile Gln Gln Leu Arg Ser									
55 60 65									
gcg cct ccc tcc ctc ggg gag ccc agc gag gcc agg tca gca tgg gca									416
Ala Pro Pro Ser Leu Gly Glu Pro Ser Glu Ala Arg Ser Ala Trp Ala									
70 75 80 85									
kag ctt tcc agt ggg aaa tgc ctc ggg cta gac gtg aga taagtcctgt									465
Xaa Leu Ser Ser Gly Lys Cys Leu Gly Leu Asp Val Arg									
90 95									
cccagtccttg actcagcaaa atcttgcttt gtggctctgt gtgagccttc ggtttc									521

<210> 445

<211> 625

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 281..547

<400> 445

atgctcgcca gaatacccoct cgttaaaggc aaggcggtt ctggctcttc cgcagctcag	60
ttatgggttc ctgtgtgctt agtcatcggg tccgacacag gctggactga tctggggagc	120
cgcgaagggc ctgccttcac aaaggacgt aacgcaagta ctgcgggcag tggttgaata	180
tggccctgaa caatgtgtcc ctgtcctccg gtgatcagag gagcagggtg gcctaccgct	240
cttcccatgg cgacctcaga ccgcgggcgt casgttggcg	295
Met Val Ser Gly Asp	
1 5	
ggc ttc ctc gtt tcc agg cct gaa gcg att cat cta gga cct cgg cag	343
Gly Phe Leu Val Ser Arg Pro Glu Ala Ile His Leu Gly Pro Arg Gln	
10 15 20	
gcg gtg cga cca agc gtt cgg gcc gag agc cgt cga gtg gat ggt ggc	391
Ala Val Arg Pro Ser Val Arg Ala Glu Ser Arg Arg Val Asp Gly Gly	
25 30 35	
ggc cgg agc cca agg gaa cca gat ggc cgg ggc cgg agc cgc caa gcs	439
Gly Arg Ser Pro Arg Glu Pro Asp Gly Arg Gly Arg Ser Arg Gln Ala	
40 45 50	
aga ttc tca cct tac cca atc cct gcc gtt gaa ccc gat ctc cta aga	487
Arg Phe Ser Pro Tyr Pro Ile Pro Ala Val Glu Pro Asp Leu Leu Arg	
55 60 65	
agt gtg ctg caa cag cgt ttg att gca tta gga ggt gtt atc gca gct	535
Ser Val Leu Gln Gln Arg Leu Ile Ala Leu Gly Gly Val Ile Ala Ala	
70 75 80 85	
cga att tca gtt taaacgaaca cctttcctct ggccctcact tagcttgtga	587
Arg Ile Ser Val	
acaggccttt ttaaaatcct ttcttggtgt agcagcaa	625

<210> 446

<211> 534

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 8..523

<400> 446

agtcaag atg gcg gga gca gct acc cag gct tcc ctg gag tcg gcc cca	49
Met Ala Gly Ala Thr Gln Ala Ser Leu Glu Ser Ala Pro	
1 5 10	
cgg atc atg cgg ctg gtg gcc gaa tgc agc cgc tcc agg gcc cgg gca	97
Arg Ile Met Arg Leu Val Ala Glu Cys Ser Arg Ser Arg Ala Arg Ala	
15 20 25 30	
ggc gag ctg tgg ctg ccg cat ggg aca gtg gcc act cct gtg ttc atg	145
Gly Glu Leu Trp Leu Pro His Gly Thr Val Ala Thr Pro Val Phe Met	
35 40 45	
cca gtg ggc acg cag gcc acc atg aag ggc atc acg acc gaa cag ctg	193
Pro Val Gly Thr Gln Ala Thr Met Lys Gly Ile Thr Thr Glu Gln Leu	
50 55 60	
gac gct ctg ggt tgc cgc atc tgc ctg ggc aat acc tac cat ctg ggt	241
Asp Ala Leu Gly Cys Arg Ile Cys Leu Gly Asn Thr Tyr His Leu Gly	
65 70 75	
cta agg ccg gga ccc gag ctg atc cag aaa gcc aac ggt ctc cac ggc	289
Leu Arg Pro Gly Pro Glu Leu Ile Gln Lys Ala Asn Gly Leu His Gly	
80 85 90	
ttc atg aat tgg cct cat aat ctg cta acg gac agc ggc ggt ttc cag	337
Phe Met Asn Trp Pro His Asn Leu Leu Thr Asp Ser Gly Gly Phe Gln	
95 100 105 110	

257

atg	gtg	tgc	ctg	gtg	tct	ctg	tcc	gag	gtg	acg	gag	gag	ggc	gtc	cgc	385
Met	Val	Ser	Leu	Val	Ser	Leu	Ser	Glu	Val	Thr	Glu	Glu	Gly	Val	Arg	
			115						120					125		
ttc	cgc	tcc	ccc	tac	gac	ggc	aat	gag	acc	ctg	ctg	agc	ccg	gag	aaa	433
Phe	Arg	Ser	Pro	Tyr	Asp	Gly	Asn	Glu	Thr	Leu	Leu	Ser	Pro	Glu	Lys	
			130					135					140			
tcc	gtg	cag	atc	cag	aat	gcg	ctg	ggc	tgc	gac	atc	atc	atg	cag	ctg	481
Ser	Val	Gln	Ile	Gln	Asn	Ala	Leu	Gly	Ser	Asp	Ile	Ile	Met	Gln	Leu	
			145				150					155				
gac	gac	gtg	gtt	agc	agt	act	gtg	act	ggg	cca	cgt	gtg	gag			523
Asp	Asp	Val	Val	Ser	Ser	Thr	Val	Thr	Gly	Pro	Arg	Val	Glu			
			160			165					170					
taggccatgt	a															534

<210> 447

<211> 111

<212> DNA

<213> Homo sapiens

<400> 447

agagctgagc	tgaagcggga	cccgagagccc	gagcagccgc	cgccatggca	atcaaatttc	60
tggaagtcac	caagcccttc	tgtgtcatcc	tgcaactaac	atctgtgaag	c	111

<210> 448

<211> 150

<212> DNA

<213> Homo sapiens

<400> 448

ttttctcaac	aattcctcac	cgcagagcct	ctgaagctcc	caccaggcca	gctctcctcc	60
accacccgct	aactctcagc	cccagtcacc	ctcttgagc	ttccctgctt	tgaattaaag	120
accactcatg	cgcaaaaaaa	aaaaaaaaaa				150

<210> 449

<211> 465

<212> DNA

<213> Homo sapiens

<400> 449

aaccaaagca	tgggctcttt	gtttcctctt	tgccaggcct	gggtccttcc	cagtgacccc	60
tcagccttca	cttctcctc	actcccggg	agctgtgaca	ggaagaggag	atcccagaac	120
tggagattga	cgtggatgag	ctcctggaca	tggagagtga	cgatgcccg	gctgccagg	180
tcaaggagct	gctggttgac	tgttacaac	ccacagaggc	cttcatttct	ggcctgctgg	240
acaagatccg	gggcatgcag	aagctgagca	caccccagaa	gaagtgagg	tccccgaccc	300
aggagaacgg	tggtccccc	aggacaawtc	gctgcccccc	aacctcgtag	caacakgcaa	360
taccggggga	ccctgcggcc	aggctggtgc	catgagcagg	gctcctcgtg	cccctggccc	420
aggggtctct	tccctgccc	cctcagtttt	ccacttttgg	ggtyt		465

<210> 450

<211> 394

<212> DNA

<213> Homo sapiens

<400> 450

agtctgaaga	tggcggcctc	agcagcgcca	ggtgctgcgg	cgctgcgtag	aagtatcaat	60
cagccggttg	cttttgtgag	agaattcct	tggactgcgg	cgctcgagtca	gctgaaagaa	120
cactttgcac	agttcgccca	tgtcagaagg	tgcattttac	cttttgacaa	ggagactggc	180
tttcacagag	gtttgggttg	ggttcagttt	tcttcagaag	aaggacttcg	gaatgcacta	240
caacaggaaa	atcatattat	agatggagta	aaggtccagg	ttcacactag	aaggccaaaa	300
cttccgcaaa	catctgatga	tgaaaagaaa	gattttttgag	actgcagcct	attaataaag	360
ttaacataac	tgacaaaaaa	aaaaaaaaaa	arsc			394

<210> 451

<211> 391
 <212> DNA
 <213> Homo sapiens

<400> 451
 aagtctgaag atggcggcct cagcagcgcg aggtgctgcg gcgctgcgta gaagtatcaa 60
 tcagccggtt gcttttgtga gaagaattcc ttggactccg gcgtcgagtc agctgaaaga 120
 acactttgca cagttcggcc atgtcagaag gtgcatttta ctttttgaca aggagactgg 180
 ctttcacaga ggtttgggtt gggttcagtt ttcttcagaa gaaggacttc ggaatgcact 240
 acaacaggaa aatcatatta tagatggagt aaaggtccag gttcacacta gaaggccaaa 300
 acttccgcaa acatctgatg atgaaaagaa agatttttga gactgcagcc tattaataaa 360
 gttaacataa ccaaaaaaaaa aaaaaaaaaag s 391

<210> 452
 <211> 427
 <212> DNA
 <213> Homo sapiens

<400> 452
 cccartcaga gggctgaaca tgtccgagtt tgtctgtaac ctgtcagcaa ggccttatgt 60
 gtacgacctc attgccgtgt ccaatcatta tggagccatg ggggttggcc actactaaag 120
 cagcttatgt gctattttac caacgtcgag atgatgaatt ttataagaca ctttcactta 180
 gcagttcttg ttctctgatg ggagggacac gaccaagcag ctctcagcag ggctttgggg 240
 atgatgaggc ttgcagcatg gacaccaact aatgctgact ccacgatcct gccaccctgt 300
 agcgccagtg taatccccc a ggagaacatc tttgacactc tgcagactgc tagtgttctg 360
 tctaaaaacc agacaaggaa atacccttct tttatgagca gaaggaaa aaaaaaaaaa 420
 aaaaagg 427

<210> 453
 <211> 438
 <212> DNA
 <213> Homo sapiens

<400> 453
 agatgttagg ctttgccaaa aactggatct ttagtatcac ttgagcactc attggacaca 60
 agaggtcaac tgttttgtca tttgtgtagt aagaacaaaa taaagtgtgt ggcattaaag 120
 aaataacact gtctggccgg gatcggtggc tcacgcctgt aatcccagca ctttgggagg 180
 ccgaggcggg cagatcacga ggtcaggagt tcgagaccag cctgaccaac atgatgaaac 240
 cctgtctcta ctaaaaatac aaaaattagc caggcatggg ggtgcacacc tgtagtccca 300
 gctgctcagg aggttgaggc aggagaatcg ctcgaaccgg ggaggtggag gttgcagtga 360
 gctgagatgg cgccactgca ttccaggcct ggggggtgac agagcaagac tcagtctcag 420
 aaaaaaaaaa aaaaaaag 438

<210> 454
 <211> 394
 <212> DNA
 <213> Homo sapiens

<400> 454
 agtctgaaga tggcggcctc agcagcgcgga ggtgctgceg cgctgcgtag aagtatcaat 60
 cagccggttg ctttttgtgag aagaattcct tggactgceg cgctcgagtc gctgaaagaa 120
 cactttgcac agttcggcca tgtcagaagg tgcattttac cttttgacaa ggagactggc 180
 tttcacagag gtttgggttg ggttcagttt tcttcagaag aaggacttcg gaatgcacta 240
 caacaggaaa atcatattat agatggagta aaggtccagg ttcacactag aaggccaaaa 300
 cttccgcaaa catctgatga tgaaaagaaa gattttttgag actgcagcct attaataaag 360
 ttaacataac tgamaaaaaa aaaaaaaaaa aggc 394

<210> 455
 <211> 713
 <212> DNA
 <213> Homo sapiens

<400> 455

259

```

atataggcgt ggctgaatac gggcctcctg cgtacgtgcg tgggtgtacgt acgtgctga      60
ttacgcgcam. acgtacgttc ctcatgaaag ggacgacggg agctgcatga aagccgaagt    120
tatggaccgc tagcatctgt cactggccac cggtttcgag gagtaagcgg cagctacctt    180
acagccctga cagcagccgg gtgctctctc ttctcaccgc ggcccacgtc tcctcgctgg    240
ctccggtggc ctgctgggtg cgcgaggagg cggaggactg tactctgagg ccaaaagcca    300
gagtcggccc tgaacgccc cgaactctcag ggtccagagg ccgtgagacc ggccggtga    360
aaggtaaaga aaccaagtgg aagagtgttt cctcctctgg ccgtaaagca ggtactctct    420
gcagcaccag ctgtccccgc cctactccgg accgccmcaa agactccatg ggatggacct    480
gagtcagccg aatcctagcc ccttcccttg ggctgctgtg ggtgctcgac atcagtgaca    540
gcaggaagca gcagaccatc aaggctacgg gaggccggg gcgcttgcca agatgaagtt    600
tggctgcctc tccttcgggc agccttatgc tggctttgtc ttaaattggaa tcaagactgt    660
ggagacgcgc tggcgctcctc tgctgagcag ccagcggaac tgtaccatcg ccg      713

```

<210> 456
 <211> 128
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -34..-1

<400> 456

Met	Ser	Leu	Arg	Leu	Asp	Thr	Thr	Pro	Ser	Cys	Asn	Ser	Ala	Arg	Pro
				-30					-25					-20	
Leu	His	Ala	Leu	Gln	Val	Leu	Leu	Leu	Leu	Ser	Leu	Leu	Leu	Thr	Ala
			-15					-10					-5		
Leu	Ala	Ser	Ser	Thr	Lys	Gly	Gln	Thr	Lys	Arg	Asn	Leu	Ala	Lys	Gly
		1				5					10				
Lys	Glu	Glu	Ser	Leu	Asp	Ser	Asp	Leu	Tyr	Ala	Glu	Leu	Arg	Cys	Met
15					20					25					30
Cys	Ile	Lys	Thr	Thr	Ser	Gly	Ile	His	Pro	Lys	Asn	Ile	Gln	Ser	Leu
				35					40					45	
Glu	Val	Ile	Gly	Lys	Gly	Thr	His	Cys	Asn	Gln	Val	Glu	Val	Ile	Ala
			50					55					60		
Thr	Leu	Lys	Asp	Gly	Arg	Lys	Ile	Cys	Leu	Asp	Pro	Asp	Ala	Pro	Arg
		65					70					75			
Ile	Lys	Lys	Ile	Val	Gln	Lys	Lys	Leu	Ala	Gly	Asp	Glu	Ser	Ala	Asp
	80					85					90				

<210> 457
 <211> 65
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -18..-1

<400> 457

Met	Lys	Ala	Ala	Val	Leu	Thr	Leu	Ala	Val	Leu	Phe	Leu	Thr	Gly	Ser
			-15					-10						-5	
Gln	Ala	Arg	His	Phe	Trp	Gln	Gln	Asp	Glu	Pro	Pro	Gln	Ser	Pro	Trp
		1				5					10				
Asp	Arg	Val	Lys	Asp	Leu	Ala	Thr	Val	Tyr	Val	Asp	Val	Leu	Lys	Asp
15					20				25						30
Ser	Gly	Arg	Asp	Tyr	Val	Ser	Gln	Phe	Glu	Gly	Ser	Ala	Trp	Glu	Asn
				35					40					45	

Ser

<210> 458
 <211> 80
 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -26..-1

<220>

<221> UNSURE

<222> 31

<223> Xaa = Ala,Val

<400> 458

Met	Gly	Arg	Ala	Met	Val	Ala	Arg	Leu	Gly	Leu	Gly	Leu	Leu	Leu	Leu	
	-25					-20					-15					
Ala	Leu	Leu	Leu	Pro	Thr	Gln	Ile	Tyr	Ser	Ser	Glu	Thr	Thr	Thr	Gly	
-10					-5					1				5		
Thr	Ser	Ser	Asn	Ser	Ser	Gln	Ser	Thr	Ser	Asn	Ser	Gly	Leu	Ala	Pro	
			10					15					20			
Asn	Pro	Thr	Asn	Ala	Thr	Thr	Lys	Xaa	Ala	Gly	Gly	Ala	Leu	Gln	Ser	
		25					30					35				
Thr	Ala	Ser	Leu	Phe	Val	Val	Ser	Leu	Ser	Leu	Leu	His	Leu	Tyr	Ser	
	40					45					50					

<210> 459

<211> 101

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -27..-1

<400> 459

Met	Ser	Ser	Ala	Ala	Gly	Phe	Cys	Ala	Ser	Arg	Pro	Gly	Leu	Leu	Phe	
	-25					-20						-15				
Leu	Gly	Leu	Leu	Leu	Leu	Pro	Leu	Val	Val	Ala	Phe	Ala	Ser	Ala	Glu	
-10						-5					1				5	
Ala	Glu	Glu	Asp	Gly	Asp	Leu	Gln	Cys	Leu	Cys	Val	Lys	Thr	Thr	Ser	
			10					15						20		
Gln	Val	Arg	Pro	Arg	His	Ile	Thr	Ser	Leu	Glu	Val	Ile	Lys	Ala	Gly	
		25					30					35				
Pro	His	Cys	Pro	Thr	Ala	Gln	Leu	Ile	Ala	Thr	Leu	Lys	Asn	Gly	Arg	
		40				45					50					
Lys	Ile	Cys	Leu	Asp	Leu	Gln	Ala	Pro	Leu	Tyr	Lys	Lys	Ile	Ile	Lys	
	55					60					65					
Lys	Leu	Leu	Glu	Ser												
70																

<210> 460

<211> 86

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -18..-1

<400> 460

Met	Ile	Pro	Ala	Val	Val	Leu	Leu	Leu	Leu	Leu	Leu	Val	Glu	Gln	Ala	
			-15					-10					-5			
Ala	Ala	Leu	Gly	Glu	Pro	Gln	Leu	Cys	Tyr	Ile	Leu	Asp	Ala	Ile	Leu	
		1				5					10					
Phe	Leu	Tyr	Gly	Ile	Val	Leu	Thr	Leu	Leu	Tyr	Cys	Arg	Leu	Lys	Ile	

261

15		20		25		30									
Gln	Val	Arg	Lys	Ala	Ala	Ile	Thr	Ser	Tyr	Glu	Lys	Ser	Asp	Gly	Val
				35					40					45	
Tyr	Thr	Gly	Leu	Ser	Thr	Arg	Asn	Gln	Glu	Thr	Tyr	Glu	Thr	Leu	Lys
			50					55					60		
His	Glu	Lys	Pro	Pro	Gln										
		65													

<210> 461
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -24..-1

<400> 461

Met	Val	Met	Gly	Leu	Gly	Val	Leu	Leu	Leu	Val	Phe	Val	Leu	Gly	Leu
				-20					-15					-10	
Gly	Leu	Thr	Pro	Pro	Thr	Leu	Ala	Gln	Asp	Asn	Ser	Arg	Tyr	Thr	His
			-5					1			5				
Phe	Leu	Thr	Gln	His	Tyr	Asp	Ala	Lys	Pro	Gln	Gly	Arg	Asp	Asp	Arg
	10				15					20					
Tyr	Cys	Glu	Ser	Ile	Met	Arg	Arg	Arg	Gly	Leu	Thr	Ser	Pro	Cys	Lys
25				30					35					40	
Asp	Ile	Asn	Thr	Phe	Ile	His	Gly	Asn	Lys	Arg	Arg	Ser	Arg	Pro	Ser
			45					50						55	
Val	Lys	Thr	Arg	Met	Glu	Thr	Leu	Thr	Glu	Lys	Thr				
			60					65							

<210> 462
 <211> 111
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -16..-1

<220>
 <221> UNSURE
 <222> 50
 <223> Xaa = Asp,Asn

<400> 462

Met	Leu	Leu	Ile	Leu	Leu	Ser	Val	Ala	Leu	Leu	Ala	Phe	Ser	Ser	Ala
	-15					-10					-5				
Gln	Asp	Leu	Asp	Glu	Asp	Val	Ser	Gln	Glu	Asp	Val	Pro	Leu	Val	Ile
1			5						10					15	
Ser	Asp	Gly	Gly	Asp	Ser	Glu	Gln	Phe	Ile	Asp	Glu	Glu	Arg	Gln	Gly
		20						25				30			
Pro	Pro	Leu	Gly	Gly	Gln	Gln	Ser	Gln	Pro	Ser	Ala	Gly	Asp	Gly	Asn
		35				40					45				
Gln	Xaa	Asp	Gly	Pro	Gln	Gln	Gly	Pro	Pro	Gln	Gln	Gly	Gly	Gln	Gln
	50					55				60					
Gln	Gln	Gly	Pro	Pro	Pro	Pro	Gln	Gly	Lys	Pro	Gln	Gly	Pro	Pro	Pro
65					70				75					80	
Gln	Gly	Gly	Arg	Pro	Gln	Gly	Pro	Pro	Gln	Gly	Gln	Ser	Pro	Gln	
				85					90					95	

<210> 463
 <211> 106

<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -16...-1

<400> 463
Met Leu Leu Ile Leu Leu Ser Val Ala Leu Leu Ala Phe Ser Ser Ala
-15 -10 -5
Gln Asp Leu Asp Glu Asp Val Ser Gln Glu Asp Val Pro Leu Val Ile
1 5 10 15
Ser Asp Gly Gly Asp Ser Glu Gln Phe Ile Asp Glu Glu Arg Gln Gly
20 25 30
Pro Pro Leu Gly Gly Gln Gln Ser Gln Pro Ser Ala Gly Asp Gly Asn
35 40 45
Gln Asp Asp Gly Pro Gln Gln Gly Pro Pro Gln Gln Gly Gly Gln Gln
50 55 60
Gln Gln Gly Pro Pro Pro Gln Gly Lys Pro Gln Gly Pro Pro Gln
65 70 75 80
Gln Gly Gly Gln Ser Cys Cys Cys Asp Lys
85 90

<210> 464
<211> 141
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -24...-1

<400> 464
Met Pro Ser Ser Val Ser Trp Gly Ile Leu Leu Leu Ala Gly Leu Cys
-20 -15 -10
Cys Leu Val Pro Val Ser Leu Ala Glu Asp Pro Gln Gly Asp Ala Ala
-5 1 5
Gln Lys Thr Asp Thr Ser His His Asp Gln Asp His Pro Thr Phe Asn
10 15 20
Lys Ile Thr Pro Asn Leu Ala Glu Phe Ala Phe Ser Leu Tyr Arg Gln
25 30 35 40
Leu Ala His Gln Ser Asn Ser Thr Asn Ile Phe Phe Ser Pro Val Ser
45 50 55
Ile Ala Thr Ala Phe Ala Met Leu Ser Leu Gly Thr Lys Ala Asp Thr
60 65 70
His Asp Glu Ile Leu Glu Gly Leu Asn Phe Asn Leu Thr Glu Ile Pro
75 80 85
Glu Ala Gln Ile His Glu Gly Phe Gln Glu Leu Leu Arg Thr Leu Asn
90 95 100
Gln Pro Asp Ser Gln Leu Gln Leu Thr Thr Gly Lys Asn
105 110 115

<210> 465
<211> 124
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -30...-1

<400> 465
Met Pro Ala Cys Arg Leu Gly Pro Leu Ala Ala Ala Leu Leu Leu Ser

263

```

-30          -25          -20          -15
Leu Leu Leu Phe Gly Phe Thr Leu Val Ser Gly Thr Gly Ala Glu Lys
          -10          -5          1
Thr Gly Val Cys Pro Glu Leu Gln Ala Asp Gln Asn Cys Thr Gln Glu
          5          10          15
Cys Val Ser Asp Ser Glu Cys Ala Asp Asn Leu Lys Cys Cys Ser Ala
          20          25          30
Gly Cys Ala Thr Phe Cys Ser Leu Pro Asn Asp Lys Glu Gly Ser Cys
35          40          45          50
Pro Gln Val Asn Ile Asn Phe Pro Gln Leu Gly Leu Cys Arg Asp Gln
          55          60          65
Cys Gln Val Asp Ser Gln Cys Pro Gly Gln Met Lys Cys Cys Arg Asn
          70          75          80
Gly Cys Gly Lys Val Ser Cys Val Thr Pro Asn Phe
          85          90

```

<210> 466

<211> 124

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -30..-1

<400> 466

```

Met Pro Ala Cys Arg Leu Gly Pro Leu Ala Ala Ala Leu Leu Leu Ser
-30          -25          -20          -15
Leu Leu Leu Phe Gly Phe Thr Leu Val Ser Gly Thr Gly Ala Glu Lys
          -10          -5          1
Thr Gly Val Cys Pro Glu Leu Gln Ala Asp Gln Asn Cys Thr Gln Glu
          5          10          15
Cys Val Ser Asp Ser Glu Cys Ala Asp Asn Ile Lys Cys Cys Ser Ala
          20          25          30
Gly Cys Ala Thr Phe Cys Ser Leu Pro Asn Asp Lys Glu Gly Ser Cys
35          40          45          50
Pro Gln Val Asn Ile Asn Phe Pro Gln Leu Gly Leu Cys Arg Asp Gln
          55          60          65
Cys Gln Val Asp Ser Gln Cys Pro Gly Gln Met Lys Cys Cys Arg Asn
          70          75          80
Gly Cys Gly Lys Val Ser Cys Val Thr Pro Asn Phe
          85          90

```

<210> 467

<211> 69

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -17..-1

<400> 467

```

Met Phe Pro Gly Phe Val Phe Leu Leu Phe Val Ile Ser Leu Ala Ala
          -15          -10          -5
Ala Ala His Leu Trp Val Leu Ala Ala Phe Met Gly Arg Ile Thr Val
          1          5          10          15
Lys Val Cys Ser Phe Thr Pro Glu Ala Ser Lys Thr Val Ser Pro Pro
          20          25          30
Glu Gly Ala Asn Asn Ser Arg Arg Thr Ala Phe Lys Ser Cys Asn Thr
          35          40          45
His His Lys Gly Leu
          50

```


<210> 468
 <211> 77
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -33..-1

<400> 468
 Met Gly Pro Val Lys Gln Leu Lys Arg Met Phe Glu Pro Thr Arg Leu
 -30 -25 -20
 Ile Ala Thr Ile Met Val Leu Leu Ser Phe Ala Leu Thr Leu Cys Ser
 -15 -10 -5
 Ala Phe Trp Trp His Asn Met Gly Leu Ala Leu Ile Phe Cys Ile Leu
 1 5 10 15
 Gln Ser Leu Ala Leu Thr Trp Tyr Ser Leu Ser Phe Ile Pro Phe Ala
 20 25 30
 Arg Asp Ala Val Lys Lys Cys Phe Ala Val Cys Leu Ala
 35 40

<210> 469
 <211> 146
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -26..-1

<400> 469
 Met Thr Met Arg Ser Leu Leu Arg Thr Pro Phe Leu Cys Gly Leu Leu
 -25 -20 -15
 Trp Ala Phe Cys Ala Pro Gly Ala Arg Ala Glu Glu Pro Ala Ala Ser
 -10 -5 1 5
 Phe Ser Gln Pro Gly Ser Met Gly Leu Asp Lys Asn Thr Val His Asp
 10 15 20
 Gln Glu His Ile Met Glu His Leu Glu Gly Val Ile Asn Lys Pro Glu
 25 30 35
 Ala Glu Met Ser Pro Gln Glu Leu Gln Leu His Tyr Phe Lys Met His
 40 45 50
 Asp Tyr Asp Gly Asn Asn Leu Leu Asp Gly Leu Glu Leu Ser Thr Ala
 55 60 65 70
 Ile Thr His Val His Lys Glu Glu Gly Ser Glu Gln Ala Pro Leu Met
 75 80 85
 Ser Glu Asp Glu Leu Ile Asn Ile Ile Asp Gly Val Leu Arg Asp Asp
 90 95 100
 Asp Lys Asn Asn Asp Gly Tyr Ile Asp Tyr Ala Glu Phe Ala Lys Ser
 105 110 115
 Leu Gln
 120

<210> 470
 <211> 62
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -46..-1

<400> 470

265

```

Met Asp Gln Leu Val Phe Lys Glu Thr Ile Trp Asn Asp Ala Phe Trp
  -45                -40                -35
Gln Asn Pro Trp Asp Gln Gly Gly Leu Ala Val Ile Ile Leu Phe Ile
-30                -25                -20                -15
Thr Ala Val Leu Leu Leu Ile Leu Phe Ala Ile Val Phe Gly Leu Leu
                -10                -5                1
Thr Ser Thr Glu Asn Thr Gln Cys Glu Ala Gly Glu Glu Glu
      5                10                15

```

<210> 471
 <211> 162
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

```

<400> 471
Met Arg Ile Ala Asn Arg Thr Arg Phe Ser Ser Pro Phe Leu Ala Arg
  -45                -40                -35
Gly Ala Gly Trp Thr His Gly Arg Gly Met Met Val Val Gly Thr Gly
-30                -25                -20
Thr Ser Leu Ala Leu Ser Ser Leu Leu Ser Leu Leu Leu Phe Ala Gly
-15                -10                -5                1
Met Gln Met Tyr Ser Arg Gln Leu Ala Ser Thr Glu Trp Leu Thr Ile
      5                10                15
Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe Val Phe Ser Leu Thr Ala
      20                25                30
Phe Asn Asn Leu Glu Asn Leu Val Phe Gly Lys Gly Phe Gln Ala Lys
      35                40                45
Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu Leu Ala Leu Phe Ala Ser
50                55                60                65
Gly Leu Ile His Arg Val Cys Val Thr Thr Cys Phe Ile Phe Ser Met
      70                75                80
Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser Ser Thr Leu Tyr Gln Ala
      85                90                95
Ala Ala Pro Val Leu Thr Pro Ala Lys Val Thr Gly Lys Ser Lys Lys
      100                105                110
Arg Asn
      115

```

<210> 472
 <211> 90
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -35..-1

```

<400> 472
Met Glu Thr Gly Ala Ser Ala Ser Ile Pro Glu Leu Ile Cys Glu Ala
-35                -30                -25                -20
Met Arg Arg Ile Trp Ser Leu Gly Leu Gly Leu Val Thr Leu Thr Ala
                -15                -10                -5
Ser Trp Ala Ala Leu Phe His Asp Gly Phe Ala Val Leu Gly Gly Asn
      1                5                10
Ile Val Ser Asp Leu Ser Thr Val Arg Phe Val Ala Gln Gln Gln His
      15                20                25
Phe Gln Leu Leu Asp Val Trp Thr Arg Asn Phe Arg Lys Pro Leu Gly
30                35                40                45
Ser Met Cys Phe Val Phe Leu Leu Leu Pro

```

266

50

55

<210> 473
<211> 61
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -25..-1

<220>
<221> UNSURE
<222> 27
<223> Xaa = Ala,Val

<400> 473
Met Arg Leu Leu Gln Leu Leu Phe Arg Ala Ser Pro Ala Thr Leu Leu
-25 -20 -15 -10
Leu Val Leu Cys Leu Gln Leu Gly Ala Asn Lys Ala Gln Asp Asn Thr
-5 1 5
Arg Lys Ile Ile Ile Lys Asn Phe Asp Ile Pro Lys Ser Val Arg Pro
10 15 20
Asn Asp Glu Xaa Leu Gln Cys Leu Gln Phe Lys Gln Asn
25 30 35

<210> 474
<211> 120
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -32..-1

<400> 474
Met Ala Ser Pro Lys Gly Phe Phe Asn Tyr Leu Thr Tyr Phe Leu Ala
-30 -25 -20
Ala Gly Ala Val Thr Leu Gly Ile Gly Phe Phe Ala Leu Ala Ser Ala
-15 -10 -5
Leu Trp Phe Leu Ile Cys Lys Arg Arg Glu Ile Phe Gln Asn Ser Lys
1 5 10 15
Phe Lys Ala Ile Asp Glu Arg Cys Arg Gln Arg Pro Ser Met Ala Lys
20 25 30
Ile Lys Ser His Ser Gln Cys Val Phe Ile Ser Arg Asn Phe His Thr
35 40 45
Gly Arg Phe Gln Leu Gln Gln Leu Lys Ile Ile Leu Lys Met Asn Pro
50 55 60
Asn Leu Gln Gln Lys Ile Ser Phe Val Ile Pro Gln Arg Pro Ala Pro
65 70 75 80
Gln Gln Ile Ala Ala Val Leu His
85

<210> 475
<211> 78
<212> PRT
<213> Homo sapiens

<220>
<221> SIGNAL
<222> -19..-1

<400> 475

267

Met Lys Phe Phe Val Phe Ala Leu Val Leu Ala Leu Met Ile Ser Met
 -15 -10 -5
 Ile Ser Ala Asp Ser His Glu Lys Arg His His Gly Tyr Arg Arg Lys
 1 5 10
 Phe His Glu Lys His His Ser Tyr His Ile Thr Leu Leu Pro Leu Phe
 15 20 25
 Glu Glu Ser Ser Lys Ser Asn Ala Asn Glu Lys His Tyr Asn Leu Leu
 30 35 40 45
 Tyr Thr Leu Cys Phe Arg Ile Leu Ala Phe Ser Ile Val Thr
 50 55

<210> 476
 <211> 87
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -20..-1

<400> 476
 Met Gln Lys Val Thr Leu Gly Leu Leu Val Phe Leu Ala Gly Phe Pro
 -20 -15 -10 -5
 Val Leu Asp Ala Asn Asp Leu Glu Asp Lys Asn Ser Pro Phe Tyr Tyr
 1 5 10
 Asp Trp His Ser Leu Gln Val Gly Gly Leu Ile Cys Ala Gly Val Leu
 15 20 25
 Cys Ala Met Gly Ile Ile Ile Val Met Ser Ala Lys Cys Lys Cys Lys
 30 35 40
 Phe Gly Gln Lys Ser Gly His His Pro Gly Glu Thr Pro Pro Leu Ile
 45 50 55 60
 Thr Pro Gly Ser Ala Gln Ser
 65

<210> 477
 <211> 105
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19..-1

<220>
 <221> UNSURE
 <222> 37
 <223> Xaa = * ,Lys

<400> 477
 Met Arg Gly Ala Thr Arg Val Ser Ile Met Leu Leu Leu Val Thr Val
 -15 -10 -5
 Ser Asp Cys Ala Val Ile Thr Gly Ala Cys Glu Arg Asp Val Gln Cys
 1 5 10
 Gly Ala Gly Thr Cys Cys Ala Ile Ser Leu Trp Leu Arg Gly Leu Arg
 15 20 25
 Met Cys Thr Pro Leu Gly Arg Xaa Gly Glu Glu Cys His Pro Gly Ser
 30 35 40 45
 His Lys Ile Pro Phe Phe Arg Lys Arg Lys His His Thr Cys Pro Cys
 50 55 60
 Leu Pro Asn Leu Leu Cys Ser Arg Phe Pro Asp Gly Arg Tyr Arg Cys
 65 70 75
 Ser Met Asp Leu Lys Asn Ile Asn Phe
 80 85

<210> 478
 <211> 102
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -36..-1

<220>
 <221> UNSURE
 <222> 31
 <223> Xaa = His,Arg

<400> 478
 Met Ser Asn Thr Thr Val Pro Asn Ala Pro Gln Ala Asn Ser Asp Ser
 -35 -30 -25
 Met Val Gly Tyr Val Leu Gly Pro Phe Phe Leu Ile Thr Leu Val Gly
 -20 -15 -10 -5
 Val Val Val Ala Val Val Met Tyr Val Gln Lys Lys Lys Arg Val Asp
 1 5 10
 Arg Leu Arg His His Leu Leu Pro Met Tyr Ser Tyr Asp Pro Ala Glu
 15 20 25
 Glu Leu Xaa Glu Ala Glu Gln Glu Leu Leu Ser Asp Met Gly Asp Pro
 30 35 40
 Lys Val Val His Gly Trp Gln Ser Gly Tyr Gln His Lys Arg Met Pro
 45 50 55 60
 Leu Leu Asp Val Lys Thr
 65

<210> 479
 <211> 137
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -18..-1

<400> 479
 Met Lys Ala Leu Ile Val Leu Gly Leu Val Leu Leu Ser Val Thr Val
 -15 -10 -5
 Gln Gly Lys Val Phe Glu Arg Cys Glu Leu Ala Arg Thr Leu Lys Arg
 1 5 10
 Leu Gly Met Asp Gly Tyr Arg Gly Ile Ser Leu Ala Asn Trp Met Cys
 15 20 25 30
 Leu Ala Lys Trp Glu Ser Gly Tyr Asn Thr Arg Ala Thr Asn Tyr Asn
 35 40 45
 Ala Gly Asp Arg Ser Thr Asp Tyr Gly Ile Phe Gln Ile Asn Ser Arg
 50 55 60
 Tyr Trp Cys Asn Asp Gly Lys Thr Pro Gly Ala Val Asn Ala Cys His
 65 70 75
 Leu Ser Cys Ser Gly Trp His Gly Glu Ile Val Val Lys Thr Glu Met
 80 85 90
 Ser Val Ser Met Phe Lys Val Val Glu Cys Asn Ser Arg Ile Phe Leu
 95 100 105 110
 Leu Gln Leu Ile Leu Ser Leu Ser His
 115

<210> 480
 <211> 101
 <212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -29..-1

<400> 480

Met	Arg	Glu	Glu	Lys	Lys	Pro	Phe	Glu	Arg	Glu	Arg	Glu	Ser	Val	Cys
				-25				-20						-15	
Val	Cys	Met	Cys	Val	Phe	Ser	Thr	Gln	Gly	Ala	Leu	Gly	Glu	Met	Ala
			-10					-5				1			
Ala	His	Phe	Ile	Asp	Glu	Lys	Leu	Arg	Pro	Ser	Glu	Gly	Asn	Gly	His
5						10					15				
Arg	Gly	Thr	Leu	Asp	Ser	Leu	Ser	Ser	Asp	Gln	Glu	Ser	Tyr	Ile	Pro
20					25				30					35	
Ser	Thr	Ala	Asp	Pro	Thr	Gln	Ala	Gly	Pro	Glu	Leu	Leu	His	Lys	Asn
				40				45					50		
Leu	Pro	Val	Thr	Ser	Arg	Ser	Gln	Pro	Leu	Pro	Ser	Asp	Leu	Ala	Ile
			55					60					65		
Pro	Ala	Ala	Ala	Leu											
			70												

<210> 481

<211> 128

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -26..-1

<400> 481

Met	Cys	Trp	Leu	Arg	Ala	Trp	Gly	Gln	Ile	Leu	Leu	Pro	Val	Phe	Leu
	-25					-20					-15				
Ser	Leu	Phe	Leu	Ile	Gln	Leu	Leu	Ile	Ser	Phe	Ser	Glu	Asn	Gly	Phe
-10					-5				1				5		
Ile	His	Ser	Pro	Arg	Asn	Asn	Gln	Lys	Pro	Arg	Asp	Gly	Asn	Glu	Glu
			10				15					20			
Glu	Cys	Ala	Val	Lys	Lys	Ser	Cys	Gln	Leu	Cys	Thr	Glu	Asp	Lys	Lys
	25						30					35			
Cys	Val	Trp	Cys	Ser	Glu	Glu	Lys	Ala	Cys	Lys	Lys	Tyr	Cys	Phe	Pro
40					45					50					
Tyr	Phe	Gly	Cys	Arg	Phe	Ser	Ser	Ile	Tyr	Trp	Leu	Asn	Cys	Lys	Val
55					60				65					70	
Asp	Met	Phe	Gly	Ile	Met	Met	Leu	Leu	Leu	Ile	Ala	Val	Leu	Ile	Thr
				75					80					85	
Gly	Phe	Val	Trp	Tyr	Cys	Cys	Ala	Tyr	His	Phe	Tyr	Leu	Gln	Asp	Ile
			90					95					100		

<210> 482

<211> 142

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -24..-1

<220>

<221> UNSURE

<222> 80

<223> Xaa = Leu,Arg,Trp

<220>

<221> UNSURE

<222> 79

<223> Xaa = Lys,Asn

<400> 482

```

Met Arg Leu Ser Trp Phe Arg Val Leu Thr Val Leu Ser Ile Cys Leu
              -20              -15              -10
Ser Ala Val Ala Thr Ala Thr Gly Ala Glu Gly Lys Arg Lys Leu Gln
              -5              1              5
Ile Gly Val Lys Lys Arg Val Asp His Cys Pro Ile Lys Ser Arg Lys
    10              15              20
Gly Asp Val Leu His Met His Tyr Thr Gly Lys Leu Glu Asp Gly Thr
25              30              35              40
Glu Phe Asp Ser Ser Leu Pro Gln Asn Gln Pro Phe Val Phe Ser Leu
              45              50              55
Gly Thr Gly Gln Val Ile Lys Gly Trp Asp Gln Gly Leu Leu Gly Met
              60              65              70
Cys Glu Gly Glu Lys Arg Xaa Xaa Val Ile Pro Ser Glu Leu Gly Tyr
    75              80              85
Gly Glu Arg Gly Ala Pro Pro Lys Ile Pro Gly Gly Ala Thr Leu Val
    90              95              100
Phe Glu Val Glu Leu Leu Lys Ile Glu Arg Arg Thr Glu Leu
105              110              115

```

<210> 483

<211> 70

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -24...-1

<400> 483

```

Met Ala Ser Ser Thr Ser Val Ser Thr Gly Gln Ile Leu Leu Thr Arg
              -20              -15              -10
Pro Ala Cys Leu Gly Ser Trp Ala Glu Ile Arg Ser Pro Val Arg Thr
              -5              1              5
Ile Ser Ile Ala Ser Asp Phe Pro Thr Ala Arg Val Ser Leu Trp Val
    10              15              20
Pro Pro Ala Pro Gly Met Val Pro Ile Lys Ile Ser Gly Cys Ala Asn
25              30              35              40
Trp Ala Phe Ser Pro Ala
              45

```

<210> 484

<211> 55

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -23...-1

<400> 484

```

Met Ala Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly Ala Leu Ser Gly
              -20              -15              -10
Ala Ala Ala Leu Gly Phe Ala Ser Tyr Gly Ala His Gly Ala Asn Ser
              -5              1              5
Gln Met Pro Thr Gly Arg Ser Cys Leu Thr Arg Pro Thr Asn Thr Thr
    10              15              20              25
Ser Tyr Thr Ala Trp Pro Cys

```

30

<210> 485
 <211> 55
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -23..-1

<400> 485
 Met Ala Gly Pro Ala Ala Ala Phe Arg Arg Leu Gly Ala Leu Ser Gly
 -20 -15 -10
 Ala Ala Ala Leu Gly Phe Ala Ser Tyr Gly Ala His Gly Ala Asn Ser
 -5 1 5
 Gln Met Pro Thr Gly Arg Ser Cys Leu Thr Arg Pro Thr Asn Thr Thr
 10 15 20 25
 Ser Tyr Thr Ala Trp Pro Cys
 30

<210> 486
 <211> 54
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -17..-1

<400> 486
 Met Phe Arg Arg Leu Thr Phe Ala Gln Leu Leu Phe Ala Thr Val Leu
 -15 -10 -5
 Gly Ile Ala Gly Gly Val Tyr Ile Phe Gln Pro Val Phe Glu Gln Tyr
 1 5 10 15
 Ala Lys Asp Gln Lys Glu Leu Lys Glu Lys Met Gln Leu Val Gln Glu
 20 25 30
 Ser Glu Glu Lys Lys Ser
 35

<210> 487
 <211> 83
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1

<400> 487
 Met Arg Leu Phe Leu Ser Leu Pro Val Leu Val Val Val Leu Ser Ile
 -15 -10 -5 1
 Val Leu Glu Gly Pro Ala Pro Ala Gln Gly Thr Pro Asp Val Ser Ser
 5 10 15
 Ala Leu Asp Lys Leu Lys Glu Phe Gly Asn Thr Leu Glu Asp Lys Ala
 20 25 30
 Arg Glu Leu Ile Ser Arg Ile Lys Gln Ser Glu Leu Ser Ala Lys Met
 35 40 45
 Arg Glu Trp Phe Ser Glu Thr Phe Gln Lys Val Lys Asp Lys Leu Lys
 50 55 60 65
 Ile Asp Ser

<210> 488

<211> 98
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -43..-1

<400> 488
 Met Phe Arg Ile Glu Gly Leu Ala Pro Lys Leu Asp Pro Glu Glu Met
 -40 -35 -30
 Lys Arg Lys Met Arg Glu Asp Val Ile Ser Ser Ile Arg Asn Phe Leu
 -25 -20 -15
 Ile Tyr Val Ala Leu Leu Arg Val Ser Glu Cys Leu Pro Gly Cys Asp
 -10 -5 1 5
 Cys Asp Thr Ser Gly Glu Leu Thr Asp Gly His Pro Leu Thr Leu Arg
 10 15 20
 Gly His Arg Gly Leu Arg Thr Glu Leu Asn Gly Ser Gly Glu Gln Gly
 25 30 35
 Gly Ser Ile Tyr Leu Lys Glu Ile Gly Gln His Met Lys Thr Gly His
 40 45 50
 His Ile
 55

<210> 489
 <211> 82
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -43..-1

<400> 489
 Met Ala Lys Tyr Gln Gly Glu Val Gln Ser Leu Lys Leu Asp Asp Asp
 -40 -35 -30
 Ser Val Ile Glu Gly Val Ser Asp Gln Val Leu Val Ala Val Val Val
 -25 -20 -15
 Ser Phe Ala Leu Ile Ala Thr Leu Val Tyr Ala Leu Phe Arg Asn Val
 -10 -5 1 5
 His Gln Asn Ile His Pro Glu Asn Gln Glu Leu Val Arg Val Leu Arg
 10 15 20
 Glu Gln Leu Gln Thr Glu Gln Asp Ala Pro Ala Asp Ser Thr Ala Val
 25 30 35
 Leu His

<210> 490
 <211> 99
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -28..-1

<400> 490
 Met Thr Ser Ala Leu Thr Gln Gly Leu Glu Arg Ile Pro Asp Gln Leu
 -25 -20 -15
 Gly Tyr Leu Val Leu Ser Glu Gly Ala Val Leu Ala Ser Ser Gly Asp
 -10 -5 1
 Leu Glu Asn Asp Glu Gln Ala Ala Ser Ala Ile Ser Glu Leu Val Ser
 5 10 15 20
 Thr Ala Cys Gly Phe Arg Leu His Arg Gly Met Asn Val Pro Phe Lys

273

25 30 35
 Arg Leu Ser Val Phe Gly Glu His Thr Leu Leu Val Thr Val Ser
 40 45 50
 Gly Gln Arg Val Phe Val Val Lys Arg Gln Asn Arg Gly Arg Glu Pro
 55 60 65
 Ile Asp Val
 70

<210> 491

<211> 186

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -15..-1

<220>

<221> UNSURE

<222> 148

<223> Xaa = Glu,Gln

<400> 491

Met His Val Met Ala Ala Ser Met Ala Arg Gly Gly Val Ser Ala Arg
 -15 -10 -5 1
 Val Leu Leu Gln Ala Ala Arg Gly Thr Trp Trp Asn Arg Pro Gly Gly
 5 10 15
 Thr Ser Gly Ser Gly Glu Gly Val Ala Leu Gly Thr Thr Arg Lys Phe
 20 25 30
 Gln Ala Thr Gly Ser Arg Pro Ala Gly Glu Glu Asp Ala Gly Gly Pro
 35 40 45
 Glu Arg Pro Gly Asp Val Val Asn Val Val Phe Val Asp Arg Ser Gly
 50 55 60 65
 Gln Arg Ile Pro Val Ser Gly Arg Val Gly Asp Asn Val Leu His Leu
 70 75 80
 Ala Gln Arg His Gly Val Asp Leu Glu Gly Ala Cys Glu Ala Ser Leu
 85 90 95
 Ala Cys Ser Thr Cys His Val Tyr Val Ser Glu Asp His Leu Asp Leu
 100 105 110
 Leu Pro Pro Pro Glu Glu Arg Glu Asp Asp Met Leu Asp Met Ala Pro
 115 120 125
 Leu Leu Gln Glu Asn Ser Arg Leu Gly Cys Gln Ile Val Leu Thr Pro
 130 135 140 145
 Glu Leu Xaa Gly Ala Glu Phe Thr Leu Pro Lys Ile Thr Arg Asn Phe
 150 155 160
 Tyr Val Asp Gly His Val Pro Lys Pro His
 165 170

<210> 492

<211> 123

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -48..-1

<400> 492

Met Glu Ala Leu Gly Lys Leu Lys Gln Phe Asp Ala Tyr Pro Lys Thr
 -45 -40 -35
 Leu Glu Asp Phe Arg Val Lys Thr Cys Gly Gly Ala Thr Val Thr Ile
 -30 -25 -20
 Val Ser Gly Leu Leu Met Leu Leu Leu Phe Leu Ser Glu Leu Gln Tyr

274

```

      -15          -10          -5
Tyr Leu Thr Thr Glu Val His Pro Glu Leu Tyr Val Asp Lys Ser Arg
1      5      10      15
Gly Asp Lys Leu Lys Ile Asn Ile Asp Val Leu Phe Pro His Met Pro
      20      25      30
Cys Ala Tyr Leu Ser Ile Asp Ala Met Asp Val Ala Gly Glu Gln Gln
      35      40      45
Leu Asp Val Glu His Asn Leu Phe Lys Gln Arg Leu Asp Lys Asp Gly
      50      55      60
Ile Pro Val Ser Ser Glu Ala Glu Arg His Asp
65      70      75

```

<210> 493
 <211> 77
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -29...-1

<220>
 <221> UNSURE
 <222> 36
 <223> Xaa = Gln,Arg

```

<400> 493
Met Ala Ala Thr Asp Phe Val Gln Glu Met Arg Ala Val Gly Glu Arg
      -25      -20      -15
Leu Leu Leu Lys Leu Gln Arg Leu Pro Gln Ala Glu Pro Val Glu Ile
      -10      -5      1
Val Ala Phe Ser Val Ile Ile Leu Phe Thr Ala Thr Val Leu Leu Leu
      5      10      15
Leu Leu Ile Ala Cys Ser Cys Cys Cys Thr His Cys Cys Cys Pro Glu
20      25      30      35
Xaa Arg Gly Arg Lys Val Gln Val Gln Pro Thr Pro Pro
      40      45

```

<210> 494
 <211> 107
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -27...-1

```

<400> 494
Met Leu Phe Glu Glu Ala Leu Pro Leu Ser Cys Ser Asp Pro Val Leu
      -25      -20      -15
Ser Thr Leu Ser Leu Val Gln Phe Ser Pro Ser Gly Arg Thr Gln Asp
      -10      -5      1      5
Leu Leu Ser Pro Gly Val Glu Asn Leu Ser Val Leu Asp Val Ser Pro
      10      15      20
Leu Gly Leu Ala Cys Cys Leu Leu Thr Leu Thr Met Ser Cys Pro Gly
      25      30      35
Pro Asp Pro Pro Glu Gly Pro Gly Thr Gln Arg Val Trp Gln Gly Ala
      40      45      50
Leu Arg Ile Leu Gln Leu Pro Gly Ala Pro Asp Gly Val Ser Pro Tyr
      55      60      65
Gln Pro Val Trp Ser Arg Thr Pro Asp Leu Lys
70      75      80

```

<210> 495
 <211> 112
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -39..-1

<220>
 <221> UNSURE
 <222> -37
 <223> Xaa = * ,Glu,Lys,Gln

<400> 495
 Met Leu Xaa His Leu Ser Ser Leu Pro Thr Gln Met Asp Tyr Lys Gly
 -35 -30 -25
 Gln Lys Leu Ala Glu Gln Met Phe Gln Gly Ile Ile Leu Phe Ser Ala
 -20 -15 -10
 Ile Val Gly Phe Ile Tyr Gly Tyr Val Ala Glu Gln Phe Gly Trp Thr
 -5 1 5
 Val Tyr Ile Val Met Ala Gly Phe Ala Phe Ser Cys Leu Leu Thr Leu
 10 15 20 25
 Pro Pro Trp Pro Ile Tyr Arg Arg His Pro Leu Lys Trp Leu Pro Val
 30 35 40
 Gln Ala Gln Thr Thr Arg Asn Gln Gly Lys Glu Lys Leu Arg Gly Met
 45 50 55
 Leu Lys Ile Ile Glu Val Phe Met Ile Gln His Leu Leu Leu Phe Leu
 60 65 70

<210> 496
 <211> 103
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -21..-1

<400> 496
 Met Ala Leu Cys Ala Leu Thr Arg Ala Leu Arg Ser Leu Asn Leu Ala
 -20 -15 -10
 Pro Pro Thr Val Ala Ala Pro Ala Pro Ser Leu Phe Pro Ala Ala Gln
 -5 1 5 10
 Met Met Asn Asn Gly Leu Leu Gln Gln Pro Ser Ala Leu Met Leu Leu
 15 20 25
 Pro Cys Arg Pro Val Leu Thr Ser Val Ala Leu Asn Ala Asn Phe Val
 30 35 40
 Ser Trp Lys Ser Arg Thr Lys Tyr Thr Ile Thr Pro Val Lys Met Arg
 45 50 55
 Lys Ser Gly Gly Arg Asp His Thr Gly Ala Gly Asn Val Arg Ser Asn
 60 65 70 75
 Ser Arg Pro Ser Ile Gln Arg
 80

<210> 497
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15..-1

<400> 497

```

Met Ala Val Leu Ala Gly Ser Leu Leu Gly Pro Thr Ser Arg Ser Ala
-15          -10          -5          1
Ala Leu Leu Gly Gly Arg Trp Leu Gln Pro Arg Ala Trp Leu Gly Phe
          5          10          15
Pro Asp Ala Trp Gly Leu Pro Thr Pro Gln Gln Ala Arg Gly Lys Ala
          20          25          30
Arg Gly Asn Glu Tyr Gln Pro Ser Asn Ile Lys Arg Lys Asn Lys His
          35          40          45
Gly Trp Val Arg Arg Leu Ser Thr Pro Ala Gly Val Gln Val Ile Leu
50          55          60          65
Arg Arg Met Leu Lys Gly Arg Lys Ser Leu Ser His
          70          75

```

<210> 498

<211> 87

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -39..-1

<400> 498

```

Met Lys Val Leu Leu Leu Thr Gly Leu Gly Ala Leu Phe Phe Ala Tyr
          -35          -30          -25
Tyr Trp Asp Asp Asn Phe Asp Pro Ala Ser Leu Gln Gly Ala Arg Val
          -20          -15          -10
Leu Leu Thr Gly Ala Asn Ala Gly Val Gly Glu Glu Leu Ala Tyr His
          -5          1          5
Tyr Ala Arg Leu Gly Ser His Leu Val Leu Thr Ala His Thr Glu Ala
10          15          20          25
Leu Leu Gln Lys Ala Arg Trp Leu Thr Leu Val Val Ser Thr Leu Gly
          30          35          40
Gly Arg Gly Lys Trp Ile Thr
          45

```

<210> 499

<211> 125

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -15..-1

<220>

<221> UNSURE

<222> 100

<223> Xaa = Asn,Thr

<400> 499

```

Met Lys Leu Leu Thr His Asn Leu Leu Ser Ser His Val Arg Gly Val
-15          -10          -5          1
Gly Ser Arg Gly Phe Pro Leu Arg Leu Gln Ala Thr Glu Val Arg Ile
          5          10          15
Cys Pro Val Glu Phe Asn Pro Asn Phe Val Ala Arg Met Ile Pro Lys
          20          25          30
Val Glu Trp Ser Ala Phe Leu Glu Ala Ala Asp Asn Leu Arg Leu Ile
          35          40          45
Gln Val Pro Lys Gly Pro Val Glu Gly Tyr Glu Glu Asn Glu Glu Phe
50          55          60          65

```

277

Leu	Arg	Thr	Met	His	His	Leu	Leu	Leu	Glu	Val	Glu	Val	Ile	Glu	Gly
				70					75					80	
Thr	Leu	Gln	Cys	Pro	Glu	Ser	Gly	Arg	Met	Phe	Pro	Ile	Ser	Arg	Gly
		85						90					95		
Ile	Pro	Xaa	Met	Leu	Leu	Ser	Glu	Glu	Glu	Thr	Glu	Ser			
		100					105					110			

<210> 500

<211> 52

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -22..-1

<220>

<221> UNSURE

<222> 26

<223> Xaa = Ile,Asn

<400> 500

Met	Lys	Ile	Ser	Val	Ala	Ala	Ile	Pro	Phe	Phe	Leu	Leu	Ile	Thr	Ile
		-20					-15					-10			
Ala	Leu	Gly	Thr	Lys	Thr	Glu	Ser	Ser	Ser	Arg	Gly	Pro	Tyr	His	Pro
	-5				1				5					10	
Ser	Glu	Cys	Cys	Phe	Thr	Tyr	Thr	Thr	Tyr	Lys	Ile	Pro	Arg	Gln	Xaa
			15						20					25	
Tyr	Gly	Leu	Leu												
			30												

<210> 501

<211> 72

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -42..-1

<400> 501

Met	Asn	Arg	Leu	Ala	Gly	Val	Gly	Trp	Arg	Val	Asp	Tyr	Thr	Leu	Ser
		-40					-35					-30			
Ser	Ser	Leu	Leu	Gln	Ser	Val	Glu	Glu	Pro	Met	Val	His	Leu	Arg	Leu
		-25				-20					-15				
Glu	Val	Ala	Ala	Ala	Pro	Gly	Thr	Pro	Ala	Gln	Pro	Val	Ala	Met	Ser
	-10				-5				1				5		
Leu	Ser	Ala	Asp	Lys	Phe	Gln	Val	Leu	Leu	Ala	Glu	Leu	Lys	Gln	Ala
		10						15					20		
Gln	Thr	Leu	Met	Ser	Ser	Leu	Gly								
		25					30								

<210> 502

<211> 84

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -20..-1

<400> 502

Met Ala Ala Ala Ala Val Pro Ser Leu Leu Leu Ser Leu Pro Pro His

278

[illegible]

```
<210> 503
<211> 54
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> SIGNAL  
<222> -19..-1
```

```

<400> 503
Met Lys Phe Asn Phe Val Leu Leu Ile Leu Ser Lys Val Leu Asp Asp
      -15                      -10                      -5
Thr Phe Gln Ser Val Lys Lys Trp Leu Asn Tyr Phe Gln Phe Thr Leu
      1                      5                      10
Arg Asn Gly Leu Met Trp Pro Gly Ala Val Ala His Ala Cys Asn Pro
      15                      20                      25
Ser Thr Gly Ser Arg Leu
30                      35

```

```
<210> 504
<211> 91
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> SIGNAL  
<222> -19..-1
```

```
<220>
<221> UNSURE
<222> 29
<223> Xaa = Met,Thr
```

<400> 504																
Met	Tyr	Gly	Lys	Ile	Ile	Phe	Val	Leu	Leu	Leu	Ser	Glu	Ile	Val	Ser	
				-15				-10						-5		
Ile	Ser	Ala	Leu	Ser	Thr	Thr	Glu	Val	Ala	Met	His	Thr	Ser	Thr	Ser	
			1				5					10				
Ser	Ser	Val	Thr	Lys	Ser	Tyr	Ile	Ser	Ser	Gln	Thr	Asn	Gly	Glu	Xaa	
	15					20					25					
Gly	Gln	Leu	Val	His	Arg	Phe	Thr	Val	Pro	Ala	Pro	Val	Val	Ile	Ile	
30					35				40					45		
Leu	Ile	Ile	Leu	Cys	Val	Met	Ala	Gly	Ile	Ile	Gly	Thr	Ile	Leu	Leu	
				50					55					60		
Ile	Ser	Tyr	Ser	Ile	Arg	Arg	Leu	Ile	Lys	Ala						
			65				70									

```
<210> 505
<211> 114
<212> PRT
<213> Homo sapiens
```

<220>

<221> SIGNAL

<222> -32..-1

<400> 505

```

Met Gln Met Tyr Ser Arg Gln Leu Ala Ser Thr Glu Trp Leu Thr Ile
      -30                -25                -20
Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe Val Phe Ser Leu Thr Ala
      -15                -10                -5
Phe Asn Asn Leu Glu Asn Leu Val Phe Gly Lys Gly Phe Gln Ala Lys
1      5      10      15
Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu Leu Ala Leu Phe Ala Ser
      20      25      30
Gly Leu Ile His Arg Val Cys Val Thr Thr Cys Phe Ile Phe Ser Met
      35      40      45
Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser Ser Thr Leu Tyr Gln Ala
      50      55      60
Ala Ala Pro Val Leu Thr Pro Ala Lys Val Thr Gly Lys Ser Lys Lys
65      70      75      80
Arg Asn

```

<210> 506

<211> 53

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -13..-1

<400> 506

```

Met Ala Val Ala Phe Val Leu Ser Leu Gly Val Ala Ala Leu Tyr Lys
      -10                -5                1
Phe Arg Val Ala Asp Gln Arg Lys Lys Ala Tyr Ala Asp Phe Tyr Arg
      5      10      15
Asn Tyr Asp Val Met Lys Asp Phe Glu Glu Met Arg Lys Ala Gly Ile
20      25      30      35
Phe Gln Ser Val Lys
      40

```

<210> 507

<211> 108

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -40..-1

<400> 507

```

Met Gln Lys Phe Arg Lys Met Ser Glu Thr His His Ser Val Ile Ser
-40                -35                -30                -25
Val Lys Ala Ser Ser Pro Trp Leu Ser Ser Ser Val Thr Ala Pro Ser
      -20                -15                -10
Met Val Ala Pro Val Thr Phe Ala Ser Ile Val Glu Glu Glu Leu Gln
      -5      1      5
Gln Glu Ala Ala Leu Ile Arg Ser Arg Glu Lys Pro Leu Ala Leu Ile
      10      15      20
Gln Ile Glu Glu His Ala Ile Gln Asp Leu Leu Val Phe Tyr Glu Ala
25      30      35      40
Phe Gly Asn Pro Glu Phe Val Ile Val Glu Arg Thr Pro Gln Gly
      45      50      55
Pro Leu Ala Val Pro Met Trp Asn Lys His Gly Cys

```


60

65

<210> 508
 <211> 127
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

<220>
 <221> UNSURE
 <222> -22
 <223> Xaa = * ,Cys

<220>
 <221> UNSURE
 <222> -39
 <223> Xaa = * ,Tyr

<220>
 <221> UNSURE
 <222> 57
 <223> Xaa = Ile,Leu

<220>
 <221> UNSURE
 <222> -43,-27,63
 <223> Xaa = Leu,Met

<220>
 <221> UNSURE
 <222> 3
 <223> Xaa = Lys,Gln

<220>
 <221> UNSURE
 <222> -42
 <223> Xaa = Lys,Thr

<400> 508
 Met Phe Thr Pro Xaa Xaa Val Lys Xaa Ala Tyr Tyr Asp Thr Glu Arg
 -45 -40 -35
 Ile Gly Val Asp Xaa Ile Met Lys Thr Xaa Phe Ser Pro Asn Arg Val
 -30 -25 -20
 Ile Gly Leu Ser Ser Asp Leu Gln Gln Val Gly Gly Ala Ser Ala Arg
 -15 -10 -5 1
 Ile Xaa Asp Ala Leu Ser Thr Val Leu Gln Tyr Ala Glu Asp Val Leu
 5 10 15
 Ser Gly Lys Val Ser Ala Asp Asn Thr Val Gly Arg Phe Leu Met Ser
 20 25 30
 Leu Val Asn Gln Val Pro Lys Ile Val Pro Asp Asp Phe Glu Thr Met
 35 40 45
 Leu Asn Ser Asn Ile Asn Asp Xaa Leu Met Val Thr Tyr Xaa Ala Asn
 50 55 60 65
 Leu Thr Gln Ser Gln Ile Ala Leu Asn Glu Lys Leu Val Asn Leu
 70 75 80

<210> 509
 <211> 85
 <212> PRT
 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -32..-1

<400> 509

```

Met Lys Ala Ala Leu Glu Asp Thr Leu Ala Glu Thr Glu Ala Arg Phe
      -30      -25      -20
Gly Ala Gln Leu Ala His Ile Gln Ala Leu Ile Ser Gly Ile Glu Ala
      -15      -10      -5
Gln Leu Gly Asp Val Arg Ala Asp Ser Glu Arg Gln Asn Gln Glu Tyr
1          5          10          15
Gln Arg Leu Met Asp Ile Lys Ser Arg Leu Glu Gln Glu Ile Ala Thr
      20      25      30
Tyr Arg Ser Leu Leu Glu Gly Gln Glu Asp His Tyr Asn Asn Leu Ser
      35      40      45
Ala Ser Lys Val Leu
      50

```

<210> 510

<211> 101

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -25..-1

<400> 510

```

Met Val Asp Arg Glu Leu Ala Asp Ile His Glu Asp Ala Lys Thr Cys
-25      -20      -15      -10
Leu Val Leu Cys Ser Arg Val Leu Ser Val Ile Ser Val Lys Glu Ile
      -5          1          5
Lys Thr Gln Leu Ser Leu Gly Arg His Pro Ile Ile Ser Asn Trp Phe
      10      15      20
Asp Tyr Ile Pro Ser Thr Arg Tyr Lys Asp Pro Cys Glu Leu Leu His
      25      30      35
Leu Cys Arg Leu Thr Ile Arg Asn Gln Leu Leu Thr Asn Asn Met Leu
40      45      50      55
Pro Asp Gly Ile Phe Ser Leu Leu Ile Pro Ala Arg Leu Gln Asn Tyr
      60      65      70
Leu Asn Leu Glu Ile
      75

```

<210> 511

<211> 68

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -47..-1

<400> 511

```

Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn Ser His Cys Phe Ala
      -45      -40      -35
Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr Ile His Ala Ser Arg
      -30      -25      -20
Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly Gly Ser Ser Ser Glu
-15      -10      -5          1
Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn Pro Gly Pro Gly Pro
      5          10          15
Arg Trp Cys Ser

```

20

<210> 512
 <211> 68
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

<220>
 <221> UNSURE
 <222> 9
 <223> Xaa = Cys,Trp

<400> 512
 Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn Ser His Cys Phe Ala
 -45 -40 -35
 Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr Ile His Ala Ser Arg
 -30 -25 -20
 Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly Gly Ser Ser Ser Glu
 -15 -10 -5 1
 Ser Ser Arg Val Ser Leu His Xaa Cys Ser Asn Pro Gly Pro Gly Pro
 5 10 15
 Arg Trp Cys Ser
 20

<210> 513
 <211> 129
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -16..-1

<400> 513
 Met Ala Val Leu Leu Leu Leu Leu Arg Ala Leu Arg Arg Gly Pro Gly
 -15 -10 -5
 Pro Gly Pro Arg Pro Leu Trp Gly Pro Gly Pro Ala Trp Ser Pro Gly
 1 5 10 15
 Phe Pro Ala Arg Pro Gly Arg Gly Arg Pro Tyr Met Ala Ser Arg Pro
 20 25 30
 Pro Gly Asp Leu Ala Glu Ala Gly Gly Arg Ala Leu Gln Ser Leu Gln
 35 40 45
 Leu Arg Leu Leu Thr Pro Thr Phe Glu Gly Ile Asn Gly Leu Leu Leu
 50 55 60
 Lys Gln His Leu Val Gln Asn Pro Val Arg Leu Trp Gln Leu Leu Gly
 65 70 75 80
 Gly Thr Phe Tyr Phe Asn Thr Ser Arg Leu Lys Gln Lys Asn Lys Glu
 85 90 95
 Lys Asp Lys Ser Lys Gly Lys Ala Pro Glu Glu Asp Glu Gly Ile Phe
 100 105 110
 Ile

<210> 514
 <211> 109
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL

<222> -33..-1

<220>

<221> UNSURE

<222> 43

<223> Xaa = Asp,Gly,Asn,Ser

<220>

<221> UNSURE

<222> 22

<223> Xaa = Glu,Gly

<220>

<221> UNSURE

<222> 42

<223> Xaa = Pro,Arg

<400> 514

Met	Ala	Ala	Ser	Gly	Ala	Pro	Arg	Ile	Leu	Val	Asp	Leu	Leu	Lys	Leu
			-30					-25					-20		
Asn	Val	Ala	Pro	Leu	Ala	Val	Phe	Gln	Met	Leu	Lys	Ser	Met	Cys	Ala
		-15					-10					-5			
Gly	Gln	Arg	Leu	Ala	Ser	Glu	Pro	Gln	Asp	Pro	Ala	Ala	Val	Ser	Leu
1				5					10					15	
Pro	Thr	Ser	Ser	Val	Pro	Xaa	Thr	Arg	Gly	Arg	Asn	Lys	Gly	Ser	Ala
			20					25					30		
Ala	Leu	Gly	Gly	Ala	Leu	Ala	Leu	Ala	Glu	Xaa	Xaa	Ser	Arg	Glu	Gly
		35					40						45		
Ser	Ser	Gln	Arg	Met	Pro	Arg	Gln	Pro	Ser	Ala	Thr	Arg	Leu	Pro	Lys
		50					55					60			
Gly	Gly	Gly	Pro	Gly	Lys	Ser	Pro	Thr	Arg	Gly	Ser	Thr			
65						70					75				

<210> 515

<211> 103

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -36..-1

<400> 515

Met	Ala	Asn	Leu	Phe	Ile	Arg	Lys	Met	Val	Asn	Pro	Leu	Leu	Tyr	Leu
	-35					-30					-25				
Ser	Arg	His	Thr	Val	Lys	Pro	Arg	Ala	Leu	Ser	Thr	Phe	Leu	Phe	Gly
-20					-15					-10				-5	
Ser	Ile	Arg	Gly	Ala	Ala	Pro	Val	Ala	Val	Glu	Pro	Gly	Ala	Ala	Val
			1				5					10			
Arg	Ser	Leu	Leu	Ser	Pro	Gly	Leu	Leu	Pro	His	Leu	Leu	Pro	Ala	Leu
	15					20					25				
Gly	Phe	Lys	Asn	Lys	Thr	Val	Leu	Lys	Lys	Arg	Cys	Lys	Asp	Cys	Tyr
30					35					40					
Leu	Val	Lys	Arg	Arg	Gly	Arg	Trp	Tyr	Val	Tyr	Cys	Lys	Thr	His	Pro
45					50				55						60
Arg	His	Lys	Gln	Arg	Gln	Met									
					65										

<210> 516

<211> 86

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -36..-1

<400> 516

```

Met Gly Pro Ala Leu Arg Ser Leu Gln Val Lys Lys Gly Thr Glu His
  -35                      -30                -25
Ala Asp Pro Leu Pro Phe Pro Ser Val Ser Leu Ser Gly Phe Thr Val
-20                      -15                -10                -5
Gly Thr Leu Ser Glu Thr Ser Thr Gly Gly Pro Ala Thr Pro Thr Trp
                      1                      5                      10
Lys Glu Cys Pro Ile Cys Lys Glu Arg Phe Pro Ala Glu Ser Asp Lys
      15                      20                25
Asp Ala Leu Glu Asp His Met Asp Gly His Phe Phe Phe Ser Thr Gln
      30                      35                40
Asp Pro Phe Thr Phe Glu
45                      50

```

<210> 517

<211> 52

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -30..-1

<400> 517

```

Met Ser Ser Cys Gly Ile Val Gly Ser Ser Val Ser Phe Gln Leu Asp
-30                      -25                -20                -15
Ala Val Lys Leu Leu Leu Lys Met Val Ser Ser Ala Thr Thr Glu Arg
                      -10                -5                      1
Cys Cys Asn Gly Ser Ala Asn Phe His Lys Asn Leu Cys Ala Thr Gly
      5                      10                15
Ile Lys Asn Phe
      20

```

<210> 518

<211> 182

<212> PRT

<213> Homo sapiens

<400> 518

```

Met Ser Ala Thr Val Val Asp Ala Val Asn Ala Ala Pro Leu Ser Gly
1                      5                      10                15
Ser Lys Glu Met Ser Leu Glu Glu Pro Lys Lys Met Thr Arg Glu Asp
      20                      25                30
Trp Arg Lys Lys Lys Glu Leu Glu Glu Gln Arg Lys Leu Gly Asn Ala
      35                      40                45
Pro Ala Glu Val Asp Glu Glu Gly Lys Asp Ile Asn Pro His Ile Pro
      50                      55                60
Gln Tyr Ile Ser Ser Val Pro Trp Tyr Ile Asp Pro Ser Lys Arg Pro
65                      70                75                80
Thr Leu Lys His Gln Arg Pro Gln Pro Glu Lys Gln Lys Gln Phe Ser
      85                      90                95
Ser Ser Gly Glu Trp Tyr Lys Arg Gly Val Lys Glu Asn Ser Ile Ile
      100                      105                110
Thr Lys Tyr Arg Lys Gly Ala Cys Glu Asn Cys Gly Ala Met Thr His
      115                      120                125
Lys Lys Lys Asp Cys Phe Glu Arg Pro Arg Arg Val Gly Ala Lys Phe
      130                      135                140
Thr Gly Thr Asn Ile Ala Pro Asp Glu His Val Gln Pro Gln Leu Met
145                      150                155                160

```

285

Phe Asp Tyr Asp Gly Lys Arg Asp Arg Trp Asn Gly Tyr Asn Pro Glu
165 170 175
Glu His Met Lys Ile Val
180

```
<210> 519
<211> 147
<212> PRT
<213> Homo sapiens
```

[illegible]

```
<210> 520
<211> 65
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -28..-1
```

```

<400> 520
Met Asn Leu Phe Arg Phe Leu Gly Asp Leu Ser His Leu Leu Ala Ile
      -25      -20      -15
Ile Leu Leu Leu Leu Lys Ile Trp Lys Ser Arg Ser Cys Ala Ala His
      -10      -5      1
Pro Gln Leu Pro Leu Ser Phe Cys Leu Ser Val Cys Leu Ser Val Ser
5      10      15      20
Leu Ser Leu Ser Val Ser Leu Ser Leu Ser Phe Ser Val Ser Lys Lys
      25      30      35
Lys

```

```
<210> 521
<211> 140
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -40..-1
```

<400> 521
Met Val Ser Ser Phe Arg Val Ser Glu Leu Gln Val Leu Leu Gly Phe

286

-40					-35						-30					-25
Ala	Gly	Arg	Asn	Lys	Ser	Gly	Arg	Lys	His	Asp	Leu	Leu	Met	Arg	Ala	
				-20					-15						-10	
Leu	His	Leu	Leu	Lys	Ser	Gly	Cys	Ser	Pro	Ala	Val	Gln	Ile	Lys	Ile	
			-5					1				5				
Arg	Glu	Leu	Tyr	Arg	Arg	Arg	Tyr	Pro	Arg	Thr	Leu	Glu	Gly	Leu	Ser	
	10					15					20					
Asp	Leu	Ser	Thr	Ile	Lys	Ser	Ser	Val	Phe	Ser	Leu	Asp	Gly	Gly	Ser	
25					30				35						40	
Ser	Pro	Val	Glu	Pro	Asp	Leu	Ala	Val	Ala	Gly	Ile	His	Ser	Leu	Pro	
				45					50					55		
Ser	Thr	Ser	Val	Thr	Pro	His	Ser	Pro	Ser	Ser	Pro	Val	Gly	Ser	Val	
			60					65					70			
Leu	Leu	Gln	Asp	Thr	Lys	Pro	Thr	Phe	Glu	Met	Gln	Gln	Pro	Ser	Pro	
		75					80					85				
Pro	Ile	Pro	Pro	Val	His	Pro	Asp	Val	Gln	Leu	Lys					
	90					95					100					

<210> 522

<211> 154

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 135

<223> Xaa = His,Asn,Pro,Thr

<220>

<221> UNSURE

<222> 137

<223> Xaa = Ile,Leu

<220>

<221> UNSURE

<222> 139

<223> Xaa = Pro,Arg

<220>

<221> UNSURE

<222> 136

<223> Xaa = Pro,Gln

<400> 522

Met	Ala	Glu	Val	Glu	Gln	Lys	Lys	Lys	Arg	Thr	Phe	Arg	Lys	Phe	Thr	
1				5					10					15		
Tyr	Arg	Gly	Val	Asp	Leu	Asp	Gln	Leu	Leu	Asp	Met	Ser	Tyr	Glu	Gln	
			20					25					30			
Leu	Met	Gln	Leu	Tyr	Ser	Ala	Arg	Gln	Ala	Ala	Ala	Glu	Pro	Gly	Pro	
		35					40					45				
Ala	Ala	Glu	Ala	Ala	Leu	Pro	Ala	Glu	Ala	Pro	Ala	Gln	Gly	Gln	Glu	
	50					55					60					
Gly	Gly	Ala	Ala	His	Gly	Glu	Ala	Gly	Ser	Gly	Glu	Asp	Ala	Pro	Ala	
65				70					75						80	
Gly	His	Asp	His	Pro	Thr	Arg	Asp	Gly	Gly	Gln	His	Gly	Gly	Arg	Leu	
			85					90						95		
Gln	Arg	Gln	Asp	Leu	Gln	Pro	Gly	Gly	Asp	Gln	Ala	Arg	Asp	Asp	Arg	
			100					105					110			
Pro	Leu	Pro	Gly	Arg	Val	Leu	His	His	Leu	Gln	Ala	Arg	Lys	Ala	Trp	
		115					120					125				
Pro	Ala	Arg	His	Arg	Gly	Xaa	Xaa	Xaa	Leu	Xaa	Leu	His	Pro	Ser	Gln	
	130					135					140					
Val	Met	Ala	Gln	Leu	Ile	Lys	Ala	His	Met							

145

150

<210> 523
 <211> 65
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 63
 <223> Xaa = * ,Ala,Glu,Gly,Ile,Lys,Leu,Pro,Gln,Arg,Ser,Thr,Val

<400> 523
 Met Ala Glu Glu Gln Gly Arg Glu Arg Asp Ser Val Pro Lys Pro Ser
 1 5 10 15
 Val Leu Phe Leu His Pro Asp Leu Gly Val Gly Gly Ala Glu Arg Leu
 20 25 30
 Val Leu Leu Pro Phe Pro Thr Glu Arg Arg Leu Leu Phe Cys Thr Ala
 35 40 45
 Pro Arg Ser Leu Asn Ser Leu Pro Leu Lys Lys Lys Lys Ile Xaa Lys
 50 55 60
 Ser
 65

<210> 524
 <211> 133
 <212> PRT
 <213> Homo sapiens

<400> 524
 Met Glu Tyr Val Leu Glu Val Lys Asn Ser Pro Arg His Leu Leu Lys
 1 5 10 15
 Gln Phe Thr Val Cys Asp Val Pro Leu Tyr Asp Ile Cys Asp Tyr Asn
 20 25 30
 Val Ser Arg Asp Arg Cys Gln Glu Leu Gly Cys Cys Phe Tyr Glu Gly
 35 40 45
 Val Cys Tyr Lys Lys Ala Val Pro Ile Tyr Ile His Val Phe Ser Ala
 50 55 60
 Leu Ile Val Ile Ile Ala Gly Ala Phe Val Ile Thr Ile Ile Tyr Arg
 65 70 75 80
 Val Ile Gln Glu Ser Arg Lys Glu Lys Ala Ile Pro Val Tyr Val Ala
 85 90 95
 Leu Pro Gln Lys Ser Ser Glu Lys Ala Glu Leu Ala Ser Ser Ser Ser
 100 105 110
 Lys Leu Gly Leu Lys Leu Arg Val Leu Gly Leu Lys Val Leu Gly Leu
 115 120 125
 Asp Glu Ser Asp Glu
 130

<210> 525
 <211> 149
 <212> PRT
 <213> Homo sapiens

<400> 525
 Met Arg Tyr Asn Glu Lys Glu Leu Gln Ala Leu Ser Arg Gln Pro Ala
 1 5 10 15
 Glu Met Ala Ala Glu Leu Gly Met Arg Gly Pro Lys Lys Gly Ser Val
 20 25 30
 Leu Lys Arg Arg Leu Val Lys Leu Val Val Asn Phe Leu Phe Tyr Phe
 35 40 45
 Arg Thr Asp Glu Ala Glu Pro Val Gly Ala Leu Leu Leu Glu Arg Cys
 50 55 60

288

Arg Val Val Arg Glu Glu Pro Gly Thr Phe Ser Ile Ser Phe Ile Glu
 65 70 75 80
 Asp Pro Glu Arg Lys Tyr His Phe Glu Cys Ser Ser Glu Glu Gln Cys
 85 90 95
 Gln Glu Trp Met Glu Ala Leu Arg Arg Ala Ser Tyr Glu Phe Met Arg
 100 105 110
 Arg Ser Leu Ile Phe Tyr Arg Asn Glu Ile Arg Lys Val Thr Gly Lys
 115 120 125
 Asp Pro Leu Glu Gln Phe Gly Ile Ser Glu Glu Ala Arg Phe Gln Leu
 130 135 140
 Ser Gly Leu Gln Ala
 145

<210> 526
 <211> 92
 <212> PRT
 <213> Homo sapiens

<400> 526
 Met Pro Val Val Pro Ala Leu Gly Arg Pro Arg Trp Ala Asp His Leu
 1 5 10 15
 Arg Ser Gly Val Arg Asp Gln Pro Gly Gln Pro Gly Glu Ala Pro Pro
 20 25 30
 Ser Leu Leu Lys Ile Gln Lys Leu Ala Gly Tyr Gly Gly Gly Cys Leu
 35 40 45
 Trp Ser Gln Leu Leu Gly Arg Leu Arg Arg Glu Asn His Leu Ser Pro
 50 55 60
 Gly Gly Gly Gly Cys Ser Glu Pro Arg Leu Cys His Cys Thr Pro Ala
 65 70 75 80
 Trp Val Thr Glu Gln Asp Ser Ile Ser Lys Ile Glu
 85 90

<210> 527
 <211> 60
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -21..-1

<220>
 <221> UNSURE
 <222> -5
 <223> Xaa = Ser,Thr

<400> 527
 Met Asp Pro Asn Cys Ser Cys Ala Ala Gly Val Ser Cys Thr Cys Ala
 -20 -15 -10
 Xaa Pro Ala Ser Ala Lys Ser Ala Asn Ala Pro Pro Ala Arg Arg Ala
 -5 1 5 10
 Ala Ala Pro Gly Ala Leu Ala Pro Pro Ser Trp Thr Pro Ala Leu Pro
 15 20 25
 Cys His Leu Ser Val Cys Pro Lys Glu Val Leu Val
 30 35

<210> 528
 <211> 120
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL

<222> -35..-1

<400> 528

```

Met Ser Lys His Leu His Cys Lys Val Leu Gly Leu Pro Leu Arg Leu
-35          -30          -25          -20
Glu Val Ser Ala Ser Cys Leu Ser Gln Ser Leu Ala Met Ser Pro Gln
          -15          -10          -5
Thr His Ser Gln Thr Cys Ile Arg Asn Leu Val Thr Cys Ile Asn Tyr
          1          5          10
Pro Arg Thr Ser Thr Gly Cys Lys Gly Thr Thr Thr Gln Arg Ile Met
          15          20          25
Glu Pro Val Glu Leu Glu Val Glu Gly Thr Glu Gln Asp Asn Ala Lys
30          35          40          45
Thr Cys Gly Ser Leu Gly Arg Gly Asn Glu Asn Thr Met Leu Arg Gly
          50          55          60
Gly Phe Ser Met Asn Thr Thr Val Gly Gln Gly Ile Ser Lys Gln Thr
          65          70          75
His His Thr Ser Thr Thr Ser Ser
          80          85

```

<210> 529

<211> 65

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -14..-1

<400> 529

```

Met Leu Cys Trp Leu Lys Leu Val Ser Ile Thr Thr Val Ala Ile Ile
          -10          -5          1
Leu Asn Trp Ala Gln His Ala Glu Asn Thr Thr Glu Cys Ala His Trp
          5          10          15
Leu Ser Leu Ile Gln Val Thr Leu Thr Ser Leu Glu Ala Cys Gln Asn
          20          25          30
Arg Leu Val Lys Ser Lys Pro Phe His Leu Gln Asn Phe Thr Cys Lys
35          40          45          50
Pro

```

<210> 530

<211> 99

<212> PRT

<213> Homo sapiens

<400> 530

```

Met Lys Lys Lys Glu Glu Thr Thr Leu Ser Glu Met Glu Pro Val Glu
1          5          10          15
Pro Gln Tyr Gln Leu Val Asn Ala Glu Ser Thr Ser Pro Phe Leu His
          20          25          30
Cys Leu Arg Glu Val Ile Gly Glu Tyr Ser Val His Glu Phe Ser Leu
          35          40          45
Leu Gly Lys Thr Glu Ser Gln Gly Ile Gly Leu Trp Ile Ala Leu Val
          50          55          60
Val Phe Leu Ser Phe Leu Ile Phe Ser Thr Ser Phe Tyr Ile Ser Asn
65          70          75          80
Ala Glu Gln Pro Phe Phe Lys Glu Pro Pro Thr Glu Ala Ala Lys Glu
          85          90          95
Leu Ser Leu

```

<210> 531

<211> 128

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 70

<223> Xaa = Pro,Thr

<400> 531

```

Met Ala Gly Asn Leu Leu Ser Gly Ala Gly Arg Arg Leu Trp Asp Trp
1          5          10          15
Val Pro Leu Ala Cys Arg Ser Phe Ser Leu Gly Val Pro Arg Leu Ile
          20          25          30
Gly Ile Arg Leu Thr Leu Pro Pro Pro Lys Val Val Asp Arg Trp Asn
          35          40          45
Glu Lys Arg Ala Met Phe Gly Val Tyr Asp Asn Ile Gly Ile Leu Gly
          50          55          60
Asn Phe Glu Lys His Xaa Lys Glu Leu Ile Arg Gly Pro Ile Trp Leu
65          70          75          80
Arg Gly Trp Lys Gly Asn Glu Leu Gln Arg Cys Ile Arg Lys Arg Lys
          85          90          95
Met Val Gly Ser Arg Met Phe Ala Asp Asp Leu His Asn Leu Asn Lys
          100          105          110
Arg Ile Arg Tyr Leu Tyr Lys His Phe Asn Arg His Gly Lys Phe Arg
          115          120          125

```

<210> 532

<211> 52

<212> PRT

<213> Homo sapiens

<400> 532

```

Met Phe Leu Thr Leu Ala Asp Thr Cys Lys Leu Arg Gly Met Ser Phe
1          5          10          15
Leu Leu Asn Val Tyr Glu Gly Glu Ala Thr Val Ser Ser Val Leu Glu
          20          25          30
Leu Leu Glu Ser Trp Ile Ile Val Gly Asn Glu Arg Tyr Phe Asp Gly
          35          40          45
Ile Ser Ser His
          50

```

<210> 533

<211> 84

<212> PRT

<213> Homo sapiens

<400> 533

```

Met Lys Val Lys Ile Lys Cys Trp Asn Gly Val Ala Thr Trp Leu Trp
1          5          10          15
Val Ala Asn Asp Glu Asn Cys Gly Ile Cys Arg Met Ala Phe Asn Gly
          20          25          30
Cys Cys Pro Asp Cys Lys Val Pro Gly Asp Asp Cys Pro Leu Val Trp
          35          40          45
Gly Gln Cys Ser His Cys Phe His Met His Cys Ile Leu Lys Trp Leu
          50          55          60
His Ala Gln Gln Val Gln Gln His Cys Pro Met Cys Arg Gln Glu Trp
65          70          75          80
Lys Phe Lys Glu

```

<210> 534

<211> 80

<212> PRT

<213> Homo sapiens

<400> 534

```

Met Ala Thr Pro Thr Gln Thr Pro Thr Lys Ala Pro Glu Glu Pro Asp
1          5          10          15
Pro Phe Tyr Tyr Asp Tyr Asn Thr Val Gln Thr Val Gly Met Thr Leu
          20          25          30
Ala Thr Ile Leu Phe Leu Leu Gly Ile Leu Ile Val Ile Ser Lys Lys
          35          40          45
Val Lys Cys Arg Lys Ala Asp Ser Arg Ser Glu Ser Pro Thr Cys Lys
          50          55          60
Ser Cys Lys Ser Glu Leu Pro Ser Ser Ala Pro Gly Gly Gly Gly Val
65          70          75          80

```

<210> 535

<211> 127

<212> PRT

<213> Homo sapiens

<400> 535

```

Met Ser Phe Ser Gly Lys Tyr Gln Leu Gln Ser Gln Glu Asn Phe Glu
1          5          10          15
Ala Phe Met Lys Ala Ile Gly Leu Pro Glu Glu Leu Ile Gln Lys Gly
          20          25          30
Lys Asp Ile Lys Gly Val Ser Glu Ile Val Gln Asn Gly Lys His Phe
          35          40          45
Lys Phe Thr Ile Thr Ala Gly Ser Lys Val Ile Gln Asn Glu Phe Thr
          50          55          60
Val Gly Glu Glu Cys Glu Leu Glu Thr Met Thr Gly Glu Lys Val Lys
65          70          75          80
Thr Val Val Gln Leu Glu Gly Asp Asn Lys Leu Val Thr Thr Phe Lys
          85          90          95
Asn Ile Lys Ser Val Thr Glu Leu Asn Gly Asp Ile Ile Thr Asn Thr
          100          105          110
Met Thr Leu Gly Asp Ile Val Phe Lys Arg Ile Ser Lys Arg Ile
          115          120          125

```

<210> 536

<211> 77

<212> PRT

<213> Homo sapiens

<400> 536

```

Met Ala Ala Gly Thr Ser Pro Ala Asp Ala Leu Cys Asp Gly Lys Lys
1          5          10          15
Ala Ala Gly Pro Thr Ser Gly Ala Phe Tyr Asn Glu Leu Arg Ala Leu
          20          25          30
Glu Ser Leu Pro Gly Asn Tyr Ser Ser Ala Ser Thr Val Val Leu Leu
          35          40          45
Trp Ile Ala Ser Arg Gln Lys Ser Gln Ser Ser Arg Asn Ser Gln Cys
          50          55          60
Ser Arg Lys Asp Lys Gly Gly Lys Gln Glu Glu His Glu
65          70          75

```

<210> 537

<211> 75

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 65

<223> Xaa = Asp, Gly

<400> 537

292

Met Pro Gly Pro Thr Pro Ser Gly Thr Asn Val Gly Ser Ser Gly Arg
 1 5 10 15
 Ser Pro Ser Lys Ala Val Ala Ala Arg Ala Arg Asp Pro Leu Ser Gly
 20 25 30
 Arg Gly Lys Met Pro Ala Val Gly Gln Gly Val Gln Ala Ala Gln Pro
 35 40 45
 Arg Gln Ala Pro Gly Gly Cys Gly Asp Ser Thr Gln Lys Ile His Leu
 50 55 60
 Xaa Ser Lys Leu Ala Leu Phe Gln Tyr Trp Leu
 65 70 75

<210> 538
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 538

Met Ser Ser Trp Ser Leu Met Cys Met Gln Asp Met Gly Asn Thr Lys
 1 5 10 15
 Ala Arg Leu Leu Tyr Glu Ala Asn Leu Pro Glu Asn Phe Arg Arg Pro
 20 25 30
 Gln Thr Asp Gln Ala Val Glu Phe Phe Ile Arg Asp Lys Tyr Glu Lys
 35 40 45
 Lys Lys Tyr Tyr Glu Lys Met Pro
 50 55

<210> 539
 <211> 122
 <212> PRT
 <213> Homo sapiens

<400> 539

Met Glu Ala Ser Ala Leu Thr Ser Ser Ala Val Thr Ser Val Ala Lys
 1 5 10 15
 Val Val Arg Val Ala Ser Gly Ser Ala Val Val Leu Pro Leu Ala Arg
 20 25 30
 Ile Ala Thr Val Val Ile Gly Gly Val Val Ala Met Ala Ala Val Pro
 35 40 45
 Met Val Leu Ser Ala Met Gly Phe Thr Ala Ala Gly Ile Ala Ser Ser
 50 55 60
 Ser Ile Ala Ala Lys Met Met Ser Ala Ala Ala Ile Ala Asn Gly Gly
 65 70 75 80
 Gly Val Ala Ser Gly Ser Leu Val Ala Thr Leu Gln Ser Leu Gly Ala
 85 90 95
 Thr Gly Leu Ser Gly Leu Thr Lys Phe Ile Leu Gly Ser Ile Gly Ser
 100 105 110
 Ala Ile Ala Ala Val Ile Ala Arg Phe Tyr
 115 120

<210> 540
 <211> 74
 <212> PRT
 <213> Homo sapiens

<400> 540

Met Ala Ser Ser Thr Val Pro Val Ser Ala Ala Gly Ser Ala Asn Glu
 1 5 10 15
 Thr Pro Glu Ile Pro Asp Asn Val Gly Asp Trp Leu Arg Gly Val Tyr
 20 25 30
 Arg Phe Ala Thr Asp Arg Asn Asp Phe Arg Arg Asn Leu Ile Leu Asn
 35 40 45
 Leu Gly Leu Phe Ala Ala Gly Val Trp Leu Ala Arg Asn Leu Ser Asp
 50 55 60

293

Ile Asp Leu Met Ala Pro Gln Pro Gly Val
65 70

<210> 541
<211> 97
<212> PRT
<213> Homo sapiens

<400> 541
Met Ser Ser Ala Pro Glu Pro Pro Thr Phe Lys Lys Glu Pro Pro Lys
1 5 10 15
Glu Lys Glu Phe Gln Ser Pro Gly Leu Arg Gly Val Arg Thr Thr Thr
20 25 30
Leu Phe Arg Ala Val Asn Pro Glu Leu Phe Ile Lys Pro Asn Lys Pro
35 40 45
Val Met Ala Phe Gly Leu Val Thr Leu Ser Leu Cys Val Ala Tyr Ile
50 55 60
Gly Tyr Leu His Ala Ile Gln Glu Asn Lys Lys Asp Leu Tyr Glu Ala
65 70 75 80
Ile Asp Ser Glu Gly His Ser Tyr Met Arg Arg Lys Thr Ser Lys Trp
85 90 95
Asp

<210> 542
<211> 98
<212> PRT
<213> Homo sapiens

<400> 542
Met Ala Ala Glu Pro Leu Thr Glu Leu Glu Glu Ser Ile Glu Asn Val
1 5 10 15
Val Thr Thr Phe Phe Thr Phe Ala Arg Gln Glu Gly Arg Lys Asp Ser
20 25 30
Leu Ser Val Asn Glu Phe Lys Glu Leu Val Thr Gln Gln Leu Pro His
35 40 45
Leu Leu Lys Asp Val Gly Ser Leu Asp Glu Lys Met Lys Ser Leu Asp
50 55 60
Val Asn Gln Asp Ser Glu Leu Lys Phe Asn Glu Tyr Trp Arg Leu Ile
65 70 75 80
Gly Glu Leu Ala Lys Glu Ile Arg Lys Lys Lys Asp Leu Lys Ile Arg
85 90 95
Lys Lys

<210> 543
<211> 115
<212> PRT
<213> Homo sapiens

<400> 543
Met Val Ala Ala Lys Lys Thr Lys Lys Ser Leu Glu Ser Ile Asn Ser
1 5 10 15
Arg Leu Gln Leu Val Met Lys Ser Gly Lys Tyr Val Leu Gly Tyr Lys
20 25 30
Gln Thr Leu Lys Met Ile Arg Gln Gly Lys Ala Lys Leu Val Ile Leu
35 40 45
Ala Asn Asn Cys Pro Ala Leu Arg Lys Ser Glu Ile Glu Tyr Tyr Ala
50 55 60
Met Leu Ala Lys Thr Gly Val His His Tyr Ser Gly Asn Asn Ile Glu
65 70 75 80
Leu Gly Thr Ala Cys Gly Lys Tyr Tyr Arg Val Cys Thr Leu Ala Ile
85 90 95
Ile Asp Pro Gly Asp Ser Asp Ile Ile Arg Ser Met Pro Glu Gln Thr
100 105 110

Gly Glu Lys
115

<210> 544
<211> 102
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> 85
<223> Xaa = Asp,Asn

<400> 544
Met Ala Gly Gln Ala Phe Arg Lys Phe Leu Pro Leu Phe Asp Arg Val
1 5 10 15
Leu Val Glu Arg Ser Ala Ala Glu Thr Val Thr Lys Gly Gly Ile Met
20 25 30
Leu Pro Glu Lys Ser Gln Gly Lys Val Leu Gln Ala Thr Val Val Ala
35 40 45
Val Gly Ser Gly Ser Lys Gly Lys Gly Glu Ile Gln Pro Val Ser
50 55 60
Val Lys Val Gly Asp Lys Val Leu Leu Pro Glu Tyr Gly Gly Thr Lys
65 70 75 80
Val Val Leu Asp Xaa Lys Asp Tyr Phe Leu Phe Arg Asp Gly Asp Ile
85 90 95
Leu Gly Lys Tyr Val Asp
100

<210> 545
<211> 102
<212> PRT
<213> Homo sapiens

<400> 545
Met Ala Gly Gln Ala Phe Arg Lys Phe Leu Pro Leu Phe Asp Arg Val
1 5 10 15
Leu Val Glu Arg Ser Ala Ala Glu Thr Val Thr Lys Gly Gly Ile Met
20 25 30
Leu Pro Glu Lys Ser Gln Gly Lys Val Leu Gln Ala Thr Val Val Ala
35 40 45
Val Gly Ser Gly Ser Lys Gly Lys Gly Glu Ile Gln Pro Val Ser
50 55 60
Val Lys Val Gly Asp Lys Val Leu Leu Pro Glu Tyr Gly Gly Thr Lys
65 70 75 80
Val Val Leu Asp Asp Lys Asp Tyr Phe Leu Phe Arg Asp Gly Asp Ile
85 90 95
Leu Gly Lys Tyr Val Asp
100

<210> 546
<211> 116
<212> PRT
<213> Homo sapiens

<400> 546
Met Ser Ala Thr Ala Ala Thr Ala Pro Pro Ala Ala Pro Ala Gly Glu
1 5 10 15
Gly Gly Pro Pro Ala Pro Pro Pro Asn Leu Thr Ser Asn Arg Arg Leu
20 25 30
Gln Gln Thr Gln Ala Gln Val Asp Glu Val Val Asp Ile Met Arg Val
35 40 45
Asn Val Asp Lys Val Leu Glu Arg Asp Gln Lys Leu Ser Glu Leu Asp

295

50 55 60
 Asp Arg Ala Asp Ala Leu Gln Ala Gly Ala Ser Gln Phe Glu Thr Ser
 65 70 75 80
 Ala Ala Lys Leu Lys Arg Lys Tyr Trp Trp Lys Asn Leu Lys Met Met
 85 90 95
 Ile Ile Leu Gly Val Ile Cys Ala Ile Ile Leu Ile Ile Ile Ile Val
 100 105 110
 Tyr Phe Ser Thr
 115

<210> 547

<211> 97

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 54

<223> Xaa = Glu,Lys

<400> 547

Met Ser Leu Ser Leu Val Phe Arg Ala Ala Ser Tyr Phe Lys Leu Val
 1 5 10 15
 Pro Phe His Ser Ser Ser Ser Asn Gln Phe Leu Gln Pro Pro Gly Trp
 20 25 30
 Val Val Leu Thr Gln Thr Leu Val Leu Leu His Phe Glu Arg Phe Ser
 35 40 45
 Tyr Gln Asn Val Pro Xaa Ser Ala Gln Gly Lys Gly Asn Leu Gln Pro
 50 55 60
 Glu Thr Asn Ile His Leu Phe His Phe Leu Thr Phe Pro Lys Gln Ile
 65 70 75 80
 Ser Arg Asn Leu Phe Asn Ser Leu Leu Cys Leu Met Cys Leu Thr Tyr
 85 90 95
 Phe

<210> 548

<211> 113

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 110

<223> Xaa = Ser,Thr

<400> 548

Met Ser Met Thr Asp Leu Leu Asn Ala Glu Asp Ile Lys Lys Ala Val
 1 5 10 15
 Gly Ala Phe Ser Ala Thr Asp Ser Phe Asp His Lys Lys Phe Phe Gln
 20 25 30
 Met Val Gly Leu Lys Lys Lys Ser Ala Asp Asp Val Lys Lys Val Phe
 35 40 45
 His Met Leu Asp Lys Asp Lys Ser Gly Phe Ile Glu Glu Asp Glu Leu
 50 55 60
 Gly Phe Ile Leu Lys Gly Phe Ser Pro Asp Ala Arg Asp Leu Ser Ala
 65 70 75 80
 Lys Glu Thr Lys Met Leu Met Ala Ala Gly Asp Lys Asp Gly Asp Gly
 85 90 95
 Lys Ile Gly Val Asp Glu Phe Ser Thr Leu Val Ala Glu Xaa Lys Lys
 100 105 110
 His

<210> 549

<211> 112
 <212> PRT
 <213> Homo sapiens

<400> 549
 Met Asp Pro Arg Lys Val Asn Glu Leu Arg Ala Phe Val Lys Met Cys
 1 5 10 15
 Lys Gln Asp Pro Ser Val Leu His Thr Glu Glu Met Arg Phe Leu Arg
 20 25 30
 Glu Trp Val Glu Ser Met Gly Gly Lys Val Pro Pro Ala Thr Gln Lys
 35 40 45
 Ala Lys Ser Glu Glu Asn Thr Lys Glu Glu Lys Pro Asp Ser Lys Lys
 50 55 60
 Val Glu Glu Asp Leu Lys Ala Asp Glu Pro Ser Ser Glu Glu Ser Asp
 65 70 75 80
 Leu Glu Ile Asp Lys Glu Gly Val Ile Glu Pro Asp Thr Asp Ala Pro
 85 90 95
 Gln Glu Met Gly Asp Glu Asn Ala Glu Ile Thr Ala Gln His Phe Leu
 100 105 110

<210> 550
 <211> 118
 <212> PRT
 <213> Homo sapiens

<400> 550
 Met Val Leu Leu Glu Ser Glu Gln Phe Leu Thr Glu Leu Thr Arg Leu
 1 5 10 15
 Phe Gln Lys Cys Arg Thr Ser Gly Ser Val Tyr Ile Thr Leu Lys Lys
 20 25 30
 Tyr Asp Gly Arg Thr Lys Pro Ile Pro Lys Lys Gly Thr Val Glu Gly
 35 40 45
 Phe Glu Pro Ala Asp Asn Lys Cys Leu Leu Arg Ala Thr Asp Gly Lys
 50 55 60
 Lys Lys Ile Ser Thr Val Val Ser Ser Lys Glu Val Asn Lys Phe Gln
 65 70 75 80
 Met Ala Tyr Ser Asn Leu Leu Arg Ala Asn Met Asp Gly Leu Lys Lys
 85 90 95
 Arg Asp Lys Lys Asn Glu Thr Lys Lys Thr Lys Ala Ala Ala Ala Ala
 100 105 110
 Ala Ala Thr Ala Ala Gln
 115

<210> 551
 <211> 109
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 40
 <223> Xaa = Gln,Arg

<220>
 <221> UNSURE
 <222> 10
 <223> Xaa = His,Arg

<220>
 <221> UNSURE
 <222> 38
 <223> Xaa = Phe,Ile

<220>

<221> UNSURE

<222> 97

<223> Xaa = Phe, Leu

<400> 551

```

Met Asp Ser Leu Pro Ser Gly Lys Ile Xaa Arg Lys Val Lys Ile Ile
1      5      10      15
Leu Gly Arg Asn Arg Lys Glu Asn Leu Glu Pro Asn Ala Glu Phe Asp
      20      25      30
Lys Arg Thr Glu Phe Xaa Thr Xaa Glu Glu Asn Arg Ile Cys Ser Ser
      35      40      45
Pro Val Gln Ser Leu Leu Asp Leu Phe Gln Thr Ser Glu Glu Lys Ser
      50      55      60
Glu Phe Leu Gly Phe Thr Ser Tyr Thr Glu Lys Ser Gly Ile Cys Asn
65      70      75      80
Val Leu Asp Ile Trp Glu Glu Glu Asn Ser Asp Asn Leu Leu Thr Ala
      85      90      95
Xaa Phe Ser Ser Pro Ser Thr Ser Thr Phe Thr Gly Phe
      100      105

```

<210> 552

<211> 103

<212> PRT

<213> Homo sapiens

<400> 552

```

Met Ser Gly Asp Gly Ala Thr Glu Gln Ala Ala Glu Tyr Val Pro Glu
1      5      10      15
Lys Val Lys Lys Ala Glu Lys Lys Leu Glu Glu Asn Pro Tyr Asp Leu
      20      25      30
Asp Ala Trp Ser Ile Leu Ile Arg Glu Ala Gln Asn Gln Pro Ile Asp
      35      40      45
Lys Ala Arg Lys Thr Tyr Glu Arg Leu Val Ala Gln Phe Pro Ser Ser
      50      55      60
Gly Arg Phe Trp Lys Leu Tyr Ile Glu Ala Glu Val Thr Ile Leu Phe
65      70      75      80
Tyr Phe Phe Leu Tyr Gln Tyr Cys Ser Ile His Cys Ser Asp Arg Lys
      85      90      95
Gln Val Arg Asn Ile Ala Asn
      100

```

<210> 553

<211> 106

<212> PRT

<213> Homo sapiens

<400> 553

```

Met Leu Lys Ser Asn Gly Glu Arg Arg Ser Arg Asn Ala Leu Pro Ala
1      5      10      15
Val Tyr Ala Arg Lys Met Ala Ala Ser Gln Gln Gln Ala Ser Ala Ala
      20      25      30
Ser Ser Ala Ala Gly Val Ser Gly Pro Ser Ser Ala Gly Gly Pro Gly
      35      40      45
Pro Gln Gln Gln Pro Gln Pro Ala Gln Leu Val Gly Pro Ala Gln
      50      55      60
Ser Gly Leu Leu Gln Gln Gln Gln Gln Asp Phe Asp Pro Val Gln Arg
65      70      75      80
Tyr Lys Met Leu Ile Pro Gln Leu Lys Glu Ser Leu Gln Val Ile Gly
      85      90      95
Leu Lys Gln Arg Glu Ala Asn Trp Ile Trp
      100      105

```

<210> 554
 <211> 86
 <212> PRT
 <213> Homo sapiens

<400> 554
 Met Ala Glu Asp Met Glu Thr Lys Ile Lys Asn Tyr Lys Thr Ala Pro
 1 5 10 15
 Phe Asp Ser Arg Phe Pro Asn Gln Asn Gln Thr Arg Asn Cys Trp Gln
 20 25 30
 Asn Tyr Leu Asp Phe His Arg Cys Gln Lys Ala Met Thr Ala Lys Gly
 35 40 45
 Gly Asp Ile Ser Val Cys Glu Trp Tyr Gln Arg Val Tyr Gln Ser Leu
 50 55 60
 Cys Pro Thr Ser Trp Val Thr Asp Trp Asp Glu Gln Arg Ala Glu Gly
 65 70 75 80
 Thr Phe Pro Gly Lys Ile
 85

<210> 555
 <211> 105
 <212> PRT
 <213> Homo sapiens

<400> 555
 Met Cys Val Gly Leu Leu Met Val Ser Ile Leu Thr Asp Ile Asn Leu
 1 5 10 15
 Ser Asn Leu Val Ala Val Gln Tyr Ser Pro Asp Tyr Cys Asn Phe Arg
 20 25 30
 Lys Arg Ser Asp Lys Asn Gln Asp Ala Ser Thr Phe Cys His Asn Cys
 35 40 45
 Asn Gln Phe His Leu Cys Leu Gln Tyr His His Lys Ile Val Leu Pro
 50 55 60
 Trp Ser Val Leu Ala Ile Leu Ser Gln Leu Phe Leu His Ile Thr Ser
 65 70 75 80
 Arg Ile Arg Pro Thr Ile Tyr Lys Ser Asn Lys Pro Lys Ser Ile Glu
 85 90 95
 Ile Leu Ile Gly Phe Val Leu Asn Leu
 100 105

<210> 556
 <211> 101
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 95
 <223> Xaa = Cys,Trp

<220>
 <221> UNSURE
 <222> 86
 <223> Xaa = Leu,Val

<400> 556
 Met Glu Arg Pro Asp Lys Ala Ala Leu Asn Ala Leu Gln Pro Pro Glu
 1 5 10 15
 Phe Arg Asn Glu Ser Ser Leu Ala Ser Thr Leu Lys Thr Leu Leu Phe
 20 25 30
 Phe Thr Ala Leu Met Ile Thr Val Pro Ile Gly Leu Tyr Phe Thr Thr
 35 40 45
 Lys Ser Tyr Ile Phe Glu Gly Ala Leu Gly Met Ser Asn Arg Asp Ser

299

50 55 60
 Tyr Phe Tyr Ala Ala Ile Val Ala Val Val Ala Val His Val Val Leu
 65 70 75 80
 Ala Leu Phe Val Tyr Xaa Ala Trp Asn Glu Gly Ser Arg Gln Xaa Arg
 85 90 95
 Glu Gly Lys Gln Asp
 100

<210> 557
 <211> 92
 <212> PRT
 <213> Homo sapiens

<400> 557
 Met Phe Arg Asp Phe Gly Arg Arg Leu Gln Arg Asp Leu Lys Arg Val
 1 5 10 15
 Val Asp Ala Arg Leu Arg Leu Ser Glu Glu Leu Ser Gly Gly Arg Ile
 20 25 30
 Lys Pro Lys Pro Val Glu Val Gln Val Val Thr His His Met Gln Arg
 35 40 45
 Tyr Ala Val Trp Phe Gly Gly Ser Met Leu Ala Ser Thr Pro Glu Phe
 50 55 60
 Phe Gln Val Cys His Thr Lys Lys Asp Tyr Glu Glu Tyr Gly Pro Ser
 65 70 75 80
 Ile Cys Arg His Asn Pro Val Phe Gly Val Met Ser
 85 90

<210> 558
 <211> 98
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 86
 <223> Xaa = * ,Glu

<220>
 <221> UNSURE
 <222> 51
 <223> Xaa = Asp,Tyr

<400> 558
 Met Ser Val Ser Phe His Thr His Thr Lys Glu Leu Trp Thr Trp Met
 1 5 10 15
 Glu Asp Leu Gln Lys Glu Met Leu Glu Asp Val Cys Ala Asp Ser Val
 20 25 30
 Asp Ala Val Gln Glu Leu Ile Lys Gln Phe Gln Gln Gln Gln Thr Ala
 35 40 45
 Thr Leu Xaa Ala Thr Leu Asn Val Ile Lys Glu Gly Glu Asp Leu Ile
 50 55 60
 Gln Gln Leu Arg Ser Ala Pro Pro Ser Leu Gly Glu Pro Ser Glu Ala
 65 70 75 80
 Arg Ser Ala Trp Ala Xaa Leu Ser Ser Gly Lys Cys Leu Gly Leu Asp
 85 90 95
 Val Arg

<210> 559
 <211> 89
 <212> PRT
 <213> Homo sapiens

<400> 559

300

Met Val Ser Gly Asp Gly Phe Leu Val Ser Arg Pro Glu Ala Ile His
 1 5 10 15
 Leu Gly Pro Arg Gln Ala Val Arg Pro Ser Val Arg Ala Glu Ser Arg
 20 25 30
 Arg Val Asp Gly Gly Gly Arg Ser Pro Arg Glu Pro Asp Gly Arg Gly
 35 40 45
 Arg Ser Arg Gln Ala Arg Phe Ser Pro Tyr Pro Ile Pro Ala Val Glu
 50 55 60
 Pro Asp Leu Leu Arg Ser Val Leu Gln Gln Arg Leu Ile Ala Leu Gly
 65 70 75 80
 Gly Val Ile Ala Ala Arg Ile Ser Val
 85

<210> 560

<211> 172

<212> PRT

<213> Homo sapiens

<400> 560

Met Ala Gly Ala Ala Thr Gln Ala Ser Leu Glu Ser Ala Pro Arg Ile
 1 5 10 15
 Met Arg Leu Val Ala Glu Cys Ser Arg Ser Arg Ala Arg Ala Gly Glu
 20 25 30
 Leu Trp Leu Pro His Gly Thr Val Ala Thr Pro Val Phe Met Pro Val
 35 40 45
 Gly Thr Gln Ala Thr Met Lys Gly Ile Thr Thr Glu Gln Leu Asp Ala
 50 55 60
 Leu Gly Cys Arg Ile Cys Leu Gly Asn Thr Tyr His Leu Gly Leu Arg
 65 70 75 80
 Pro Gly Pro Glu Leu Ile Gln Lys Ala Asn Gly Leu His Gly Phe Met
 85 90 95
 Asn Trp Pro His Asn Leu Leu Thr Asp Ser Gly Gly Phe Gln Met Val
 100 105 110
 Ser Leu Val Ser Leu Ser Glu Val Thr Glu Glu Gly Val Arg Phe Arg
 115 120 125
 Ser Pro Tyr Asp Gly Asn Glu Thr Leu Leu Ser Pro Glu Lys Ser Val
 130 135 140
 Gln Ile Gln Asn Ala Leu Gly Ser Asp Ile Ile Met Gln Leu Asp Asp
 145 150 155 160
 Val Val Ser Ser Thr Val Thr Gly Pro Arg Val Glu
 165 170

<210> 561

<211> 528

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 57..527

<220>

<221> sig_peptide

<222> 57..143

<223> Von Heijne matrix

score 14.3000001907349

seq LLVCCLLTTPGVQG/QE

<400> 561

cttttttccc ctgcctgccc ttcgggcacc tcaggaaggc accttcctct gtcaga atg 59
 Met
 gct acc atg gta cca tcc gtg ttg tgg ccc agg gcc tgc tgg act ctg 107
 Ala Thr Met Val Pro Ser Val Leu Trp Pro Arg Ala Cys Trp Thr Leu

301

			-25					-20				-15					
ctg	gtc	tgc	tgt	ctg	ctg	acc	cca	ggt	gtc	cag	ggg	cag	gag	ttc	ctt		155
Leu	Val	Cys	Cys	Leu	Leu	Thr	Pro	Gly	Val	Gln	Gly	Gln	Glu	Phe	Leu		
		-10					-5					1					
ttg	cgg	gtg	gag	ccc	cag	aac	cct	gtg	ctc	tct	gct	gga	ggg	tcc	ctg		203
Leu	Arg	Val	Glu	Pro	Gln	Asn	Pro	Val	Leu	Ser	Ala	Gly	Gly	Ser	Leu		
5					10				15					20			
ttt	gtg	aac	tgc	agt	act	gat	tgt	ccc	agc	tct	gag	aaa	atc	gcc	ttg		251
Phe	Val	Asn	Cys	Ser	Thr	Asp	Cys	Pro	Ser	Ser	Glu	Lys	Ile	Ala	Leu		
			25					30					35				
gag	acg	tcc	cta	tca	aag	gag	ctg	gtg	gcc	agt	ggc	atg	ggc	tgg	gca		299
Glu	Thr	Ser	Leu	Ser	Lys	Glu	Leu	Val	Ala	Ser	Gly	Met	Gly	Trp	Ala		
		40					45				50						
gcc	ttc	aat	ctc	agc	aac	gtg	act	ggc	aac	agt	cgg	atc	ctc	tgc	tca		347
Ala	Phe	Asn	Leu	Ser	Asn	Val	Thr	Gly	Asn	Ser	Arg	Ile	Leu	Cys	Ser		
		55				60					65						
gtg	tac	tgc	aat	ggc	tcc	cag	ata	aca	ggc	tcc	tct	aac	atc	acc	gtg		395
Val	Tyr	Cys	Asn	Gly	Ser	Gln	Ile	Thr	Gly	Ser	Ser	Asn	Ile	Thr	Val		
	70				75					80							
tac	agg	ctc	ccg	gag	cgt	gtg	gag	ctg	gca	ccc	ctg	cct	cct	tgg	cag		443
Tyr	Arg	Leu	Pro	Glu	Arg	Val	Glu	Leu	Ala	Pro	Leu	Pro	Pro	Trp	Gln		
85				90				95						100			
ccg	gtg	ggc	cag	aac	ttc	acc	cct	gcg	ctg	cca	agt	gga	gga	tgg	gtc		491
Pro	Val	Gly	Gln	Asn	Phe	Thr	Pro	Ala	Leu	Pro	Ser	Gly	Gly	Trp	Val		
			105				110						115				
gcc	ccg	gac	cak	cct	cac	ggt	ggt	gct	gct	tcg	ctg	g					528
Ala	Pro	Asp	Xaa	Pro	His	Gly	Gly	Ala	Ala	Ser	Leu						
			120				125										

<210> 562

<211> 682

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 230..682

<220>

<221> sig_peptide

<222> 230..277

<223> Von Heijne matrix

score 13

seq LLLLLSALAGLFG/AA

<400> 562

gtgtttcgat	acagactaag	cttttaggcc	aaccctcctg	actggatggg	ggcggcgggc		60
gtggcatgca	tgaaaagtaa	acatcagaga	cctgaagaag	cttataaaat	agcttgggag		120
aggccagtca	ccaagacagg	catctcaaat	cggctgattc	tgcattctgga	aactgccttc		180
atcttgaaag	aaaagctcca	ggtcccttct	ccagccaccc	agccccaag	atg gtg atg		238
					Met Val Met		

-15

ctg	ctg	ctg	ctg	ctt	tcc	gca	ctg	gct	ggc	ctc	ttc	ggt	gcg	gca	gag		286
Leu	Leu	Leu	Leu	Leu	Ser	Ala	Leu	Ala	Gly	Leu	Phe	Gly	Ala	Ala	Glu		
			-10				-5				1						
gga	caa	gca	ttt	cat	ctt	ggg	aag	tgc	ccc	aat	cct	ccg	gtg	cag	gag		334
Gly	Gln	Ala	Phe	His	Leu	Gly	Lys	Cys	Pro	Asn	Pro	Pro	Val	Gln	Glu		
5				10				15									
aat	ttt	gac	gtg	aat	aag	tat	ctc	gga	aga	tgg	tac	gaa	att	gag	aag		382
Asn	Phe	Asp	Val	Asn	Lys	Tyr	Leu	Gly	Arg	Trp	Tyr	Glu	Ile	Glu	Lys		
20				25				30					35				
atc	cca	aca	acc	ttt	gag	aat	gga	cgc	tgc	atc	cag	gcc	aac	tac	tca		430
Ile	Pro	Thr	Thr	Phe	Glu	Asn	Gly	Arg	Cys	Ile	Gln	Ala	Asn	Tyr	Ser		

302

40	45	50	
cta atg gaa aac gga aag atc aaa gtg tta aac cag gag ttg aga gct			478
Leu Met Glu Asn Gly Lys Ile Lys Val Leu Asn Gln Glu Leu Arg Ala			
55	60	65	
gat gga act gtg aat caa atc gaa ggt gaa gcc acc cca gtt aac ctc			526
Asp Gly Thr Val Asn Gln Ile Glu Gly Glu Ala Thr Pro Val Asn Leu			
70	75	80	
aca gag cct gcc aag ctg gaa gtt aag ttt tcc tgg ttt atg cca tcg			574
Thr Glu Pro Ala Lys Leu Glu Val Lys Phe Ser Trp Phe Met Pro Ser			
85	90	95	
gca ccg tac tgg atc ctg gcc acc gac tat gag aac tat gcc ctc gtg			622
Ala Pro Tyr Trp Ile Leu Ala Thr Asp Tyr Glu Asn Tyr Ala Leu Val			
100	105	110	115
tat tcc tgt acc tgc atc atc caa ctt ttt cac gtg gat ttt gct tgg			670
Tyr Ser Cys Thr Cys Ile Ile Gln Leu Phe His Val Asp Phe Ala Trp			
120	125	130	
atc ttg gca aga			682
Ile Leu Ala Arg			
135			

<210> 563
 <211> 500
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 72..500

<220>
 <221> sig_peptide
 <222> 72..122
 <223> Von Heijne matrix
 score 10.6999998092651
 seq VLMLLAVLIWTGA/EN

<220>
 <221> misc_feature
 <222> 1..3
 <223> n=a, g, c or t

<400> 563

nnngaacatc taatcgaaaa cttgtttcct gagttgtcct gtgctggagg tctgctcaga	60
cgaaggtctc c atg gcg tta gaa gtc ttg atg ctc ctc gct gtc ttg att	110
Met Ala Leu Glu Val Leu Met Leu Leu Ala Val Leu Ile	
-15 -10 -5	
tgg acc ggt gct gag aac ctc cat gtg aaa ata agt tgc tct ctg gac	158
Trp Thr Gly Ala Glu Asn Leu His Val Lys Ile Ser Cys Ser Leu Asp	
1 5 10	
tgg ttg atg gtc tca gtt atc cca gtt gca gaa agc aga aat ctg tat	206
Trp Leu Met Val Ser Val Ile Pro Val Ala Glu Ser Arg Asn Leu Tyr	
15 20 25	
ata ttt gcg gat gaa tta cat ctg gga atg ggc tgc cct gca aat cgg	254
Ile Phe Ala Asp Glu Leu His Leu Gly Met Gly Cys Pro Ala Asn Arg	
30 35 40	
ata cat aca tat gta tat gag ttt ata tat ctt gtt cgt gat tgt ggc	302
Ile His Thr Tyr Val Tyr Glu Phe Ile Tyr Leu Val Arg Asp Cys Gly	
45 50 55 60	
atc agg aca agg gta gtt tct gag gaa act ctc ctt ttt caa acc gag	350
Ile Arg Thr Arg Val Val Ser Glu Glu Thr Leu Leu Phe Gln Thr Glu	
65 70 75	
ctg tac ttt acc cca agg aat ata gat cat gac cct cag gaa atc cat	398
Leu Tyr Phe Thr Pro Arg Asn Ile Asp His Asp Pro Gln Glu Ile His	

303

	80		85		90		
ttg gag tgt tcc acc tct agg aaa tca gtg tgg ctt aca cca gtt tct							446
Leu Glu Cys Ser Thr Ser Arg Lys Ser Val Trp Leu Thr Pro Val Ser							
	95		100		105		
act gag aat gaa ata aaa ttg gat cct agt cct ttt att gct gac ttt							494
Thr Glu Asn Glu Ile Lys Leu Asp Pro Ser Pro Phe Ile Ala Asp Phe							
	110		115		120		
cag aca							500
Gln Thr							
125							

<210> 564
 <211> 497
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 52..495

<220>
 <221> sig_peptide
 <222> 52..102
 <223> Von Heijne matrix
 score 10.6999998092651
 seq VLMLLAVLIWTGA/EN

<400> 564	
cttgtttcct gagttgtcct gtgctggagg tctgctcaga cgaagggtctc c atg gcg	57
	Met Ala
tta gaa gtc ttg atg ctc ctc gct gtc ttg att tgg acc ggt gct gag	105
Leu Glu Val Leu Met Leu Leu Ala Val Leu Ile Trp Thr Gly Ala Glu	
-15 -10 -5 1	
aac ctc cat gtg aaa ata agt tgc tct ctg gac tgg ttg atg gtc tca	153
Asn Leu His Val Lys Ile Ser Cys Ser Leu Asp Trp Leu Met Val Ser	
5 10 15	
gtt atc cca gtt gca gaa agc aga aat ctg tat ata ttt gcg gat gaa	201
Val Ile Pro Val Ala Glu Ser Arg Asn Leu Tyr Ile Phe Ala Asp Glu	
20 25 30	
tta cat ctg gga atg ggc tgc cct gca aat cgg ata cat aca tat gta	249
Leu His Leu Gly Met Gly Cys Pro Ala Asn Arg Ile His Thr Tyr Val	
35 40 45	
tat gag ttt ata tat ctt gtt cgt gat tgt ggc atc agg aca agg gta	297
Tyr Glu Phe Ile Tyr Leu Val Arg Asp Cys Gly Ile Arg Thr Arg Val	
50 55 60 65	
gtt tct gag gaa act ctc ctt ttt caa acc gag ctg tac ttt acc cca	345
Val Ser Glu Glu Thr Leu Leu Phe Gln Thr Glu Leu Tyr Phe Thr Pro	
70 75 80	
agg aat ata gat cat gac cct cag gaa atc cat ttg gag tgt tcc acc	393
Arg Asn Ile Asp His Asp Pro Gln Glu Ile His Leu Glu Cys Ser Thr	
85 90 95	
tct agg aaa tca gtg tgg ctt aca cca gtt tct act gag aat gaa ata	441
Ser Arg Lys Ser Val Trp Leu Thr Pro Val Ser Thr Glu Asn Glu Ile	
100 105 110	
aaa ttg gat cct agt cct ttt att gct gac ttt cag aca aca gca gaa	489
Lys Leu Asp Pro Ser Pro Phe Ile Ala Asp Phe Gln Thr Thr Ala Glu	
115 120 125	
gag tta gg	497
Glu Leu	
130	

<210> 565
 <211> 342

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 118..342

<220>
<221> sig_peptide
<222> 118..189
<223> Von Heijne matrix
score 10.6000003814697
seq LLLLLLLCPAHVGG/LW

<400> 565
tccccggccg ccgccgttgc gctcgccgcg ctgcgactga agcccggggc ctgcgcgcgc 60
gcggttcgcc ccgcagcctc gccccctgcc caccggggcg gccgtagggc ggtcacg 117
atg ctg ccg ccc tta ccc tcc cgc ctc ggg ctg ctg ctg ctg ctg ctc 165
Met Leu Pro Pro Leu Pro Ser Arg Leu Gly Leu Leu Leu Leu Leu
-20 -15 -10
ctg tgc ccg gcg cac gtc ggc gga ctg tgg tgg gct gtg ggc agc ccc 213
Leu Cys Pro Ala His Val Gly Gly Leu Trp Trp Ala Val Gly Ser Pro
-5 1 5
ttg gtt atg gac cct acc agc atc tgc agg aag gca cgg cgg ctg gcc 261
Leu Val Met Asp Pro Thr Ser Ile Cys Arg Lys Ala Arg Arg Leu Ala
10 15 20
ggg cgg cag gcc gag ttg tgc cag gct gag ccg gaa gtg gtg gca gag 309
Gly Arg Gln Ala Glu Leu Cys Gln Ala Glu Pro Glu Val Val Ala Glu
25 30 35 40
ctg gct cgg ggc gcc cgg ctc ggg gtg cga rag 342
Leu Ala Arg Gly Ala Arg Leu Gly Val Arg Xaa
45 50

<210> 566
<211> 434
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 242..433

<220>
<221> sig_peptide
<222> 242..307
<223> Von Heijne matrix
score 10.1999998092651
seq LLALAALLQGAVS/LK

<400> 566
acatttgccc cagggaaggt cacagctgcc tgaactbbtc aaaactccca gacacgcact 60
gcctgtgcag gatccggagc ccagcagcac tgccagggcc ttgaagtgtc tattcagaga 120
cctttcttca tagactactt ttttttcttt aagcagcaaa aggagaaaat tgatcatcaaa 180
ggatattcca gattcttgac agcattctcg tcatctctga ggacatcacc atcatctcag 240
g atg agg ggc atg aag ctg ctg ggg gcg ctg ctg gca ctg gcg gcc cta 289
Met Arg Gly Met Lys Leu Leu Gly Ala Leu Leu Ala Leu Ala Leu
-20 -15 -10
ctg cag ggg gcc gtg tcc ctg aag atc gca gcc ttc aac atc cag aca 337
Leu Gln Gly Ala Val Ser Leu Lys Ile Ala Ala Phe Asn Ile Gln Thr
-5 1 5 10
ttt ggg gag acc aag atg tcc aat gcc acc ctc gtc agc tac att gtg 385
Phe Gly Glu Thr Lys Met Ser Asn Ala Thr Leu Val Ser Tyr Ile Val
15 20 25

305

cag atc ctg agc cgc tat gac atc gcc ctg gtc cag gag gtc aga gac a 434
 Gln Ile Leu Ser Arg Tyr Asp Ile Ala Leu Val Gln Glu Val Arg Asp
 30 35 40

<210> 567
 <211> 487
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 136..486

<220>
 <221> sig_peptide
 <222> 136..234
 <223> Von Heijne matrix
 score 8.5
 seq FFFVLGSLIFCFG/IW

<400> 567
 ctcttttggg gttcttcctt tctctctcag ctctccgtct ctctttctct ctcagcctct 60
 ttctttctcc ctgtctcccc cactgtcagc acctcttctg tgtggtgagt ggaccgctta 120
 cccactagg tgaag atg tca gcc cag gag agc tgc ctc agc ctc atc aag 171
 Met Ser Ala Gln Glu Ser Cys Leu Ser Leu Ile Lys
 -30 -25
 tac ttc ctc ttc gtt ttc aac ctc ttc ttc ttc gtc ctc ggc agc ctg 219
 Tyr Phe Leu Phe Val Phe Asn Leu Phe Phe Phe Val Leu Gly Ser Leu
 -20 -15 -10
 atc ttc tgc ttc ggc atc tgg atc ctc att gac aag acc agc ttc gtg 267
 Ile Phe Cys Phe Gly Ile Trp Ile Leu Ile Asp Lys Thr Ser Phe Val
 -5 1 5 10
 tcc ttt gtg ggc ttg gcc ttc gtg cct ctg cag atc tgg tcc aaa gtc 315
 Ser Phe Val Gly Leu Ala Phe Val Pro Leu Gln Ile Trp Ser Lys Val
 15 20 25
 ctg gcc atc tca gga atc ttc acc atg ggc atc gcc ctc ctg ggt tgt 363
 Leu Ala Ile Ser Gly Ile Phe Thr Met Gly Ile Ala Leu Leu Gly Cys
 30 35 40
 gtg ggg gcc tca agg agc tcc gct gcc tcc tgg gcc tgt att ttg gga 411
 Val Gly Ala Ser Arg Ser Ser Ala Ala Ser Trp Ala Cys Ile Leu Gly
 45 50 55
 tgc tgc tgc tcc tgt ttg cca cac aga tca ccc tgg gaa tcc tca tct 459
 Cys Cys Cys Ser Cys Leu Pro His Arg Ser Pro Trp Glu Ser Ser Ser
 60 65 70 75
 cca ctc agc ggg cca gct gra gcg aag c 487
 Pro Leu Ser Gly Pro Ala Xaa Ala Lys
 80

<210> 568
 <211> 478
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 233..478

<220>
 <221> sig_peptide
 <222> 233..328
 <223> Von Heijne matrix
 score 8.19999980926514
 seq LVAAMVLLSVVFC/LY

<400> 568
 atgaatgtca gaataggccc attctaccag ctctggctga gcctgagctt ccaaaagtga 60
 gctgagctgt tcaaccttgg atcttaatta ctcttagcag ggataattag gtcctctctt 120
 ctcagattac aggtttgaga agttccaaat atccgatgcc ctagagggtt ttttccctta 180
 attttattgt tacaattggc cttcatctag ctctggatgg ttgaactgta gc atg gca 238
 Met Ala
 aag atg ttt gat ctc agg acg aag atc atg atc ggc atc gga agc agc 286
 Lys Met Phe Asp Leu Arg Thr Lys Ile Met Ile Gly Ile Gly Ser Ser
 -30 -25 -20 -15
 tta ctg gtt gcc gcg atg gtg ctc cta agt gtt gtg ttc tgt ctt tac 334
 Leu Leu Val Ala Ala Met Val Leu Leu Ser Val Val Phe Cys Leu Tyr
 -10 -5 1
 ttc aaa gta gct aag gca cta aaa gct gca aag gac cct gat gct gtg 382
 Phe Lys Val Ala Lys Ala Leu Lys Ala Ala Lys Asp Pro Asp Ala Val
 5 10 15
 gct gta aaa aat cac aac cca gac aag gtg tgt tgg gcc acg aac agc 430
 Ala Val Lys Asn His Asn Pro Asp Lys Val Cys Trp Ala Thr Asn Ser
 20 25 30
 cak gcc aaa gcc acc acc atg gag tct tgt cca tct ctc cag tgc tgt 478
 Xaa Ala Lys Ala Thr Thr Met Glu Ser Cys Pro Ser Leu Gln Cys Cys
 35 40 45 50

<210> 569
 <211> 469
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 290..469

<220>
 <221> sig_peptide
 <222> 290..346
 <223> Von Heijne matrix
 score 7.5
 seq LAVFFLVSTQLFA/EE

<400> 569
 agatgttatg ggcatcgacg agttaccgtc tcggaaactc tcaatcacgc aagcgaaagg 60
 agaggaggcg gctaattaaa tattgagcag aaagtcgcgt ggggagaatg tcacgtgggt 120
 ctggaggctc aaggaggctg ggataaatac cgcaagcact gagcaggcga aagagcgcg 180
 tcggacctcc ttcccggcgg cagctaccga gagtgcggas gaccagcgtg cgctcggagg 240
 aaccagagaa actcagcacc ccgcgggact gtcggtcgca aaatccaac atg aaa atc 298
 Met Lys Ile
 ctc gtg gcc ttg gca gtc ttt ttt ctt gtc tcc act cag ctg ttt gca 346
 Leu Val Ala Leu Ala Val Phe Phe Leu Val Ser Thr Gln Leu Phe Ala
 -15 -10 -5
 gaa gaa ata gga gcc aat gat gat ctg aat tac tgg tcc gac tgg tac 394
 Glu Glu Ile Gly Ala Asn Asp Asp Leu Asn Tyr Trp Ser Asp Trp Tyr
 1 5 10 15
 gac agc gac cag atc aag gag gaa ctg ccg gag ccc ttt gag cat ctt 442
 Asp Ser Asp Gln Ile Lys Glu Glu Leu Pro Glu Pro Phe Glu His Leu
 20 25 30
 ctg cag aga atc gcc cgg aga ccc aag 469
 Leu Gln Arg Ile Ala Arg Arg Pro Lys
 35 40

<210> 570
 <211> 489
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 253..489

<220>
 <221> sig_peptide
 <222> 253..336
 <223> Von Heijne matrix
 score 7.40000009536743
 seq CLLLNLAAPLNA/DS

<400> 570
 ctctttgctc taacagacag cagcgacttt aggctggata atagtcaa at tcttacctcg 60
 ctcttttcaact gctagtaaga tcagattgcg tttcttttcag ttactcttca atcgccagtt 120
 tcttgatctg cttctaaaag aagaagtaga gaagataaat cctgtcttca atacctggaa 180
 ggaaaaacaa aataacctca actccgtttt gaaaaaaaca ttccaagaac tttcatcaga 240
 gattttactt ag atg att tac aca atg aag aaa gta cat gca ctt tgg gct 291
 Met Ile Tyr Thr Met Lys Lys Val His Ala Leu Trp Ala
 -25 -20
 tct gta tgc ctg ctg ctt aat ctt gcc cct gcc cct ctt aat gct gat 339
 Ser Val Cys Leu Leu Leu Asn Leu Ala Pro Ala Pro Leu Asn Ala Asp
 -15 -10 -5 1
 tct gag gaa gat gaa gaa cac aca att atc aca gat acg gag ttg cca 387
 Ser Glu Glu Asp Glu Glu His Thr Ile Ile Thr Asp Thr Glu Leu Pro
 5 10 15
 cca ctg aaa ctt atg cat tca ttt tgt gca ttc aag gcg gat gat ggc 435
 Pro Leu Lys Leu Met His Ser Phe Cys Ala Phe Lys Ala Asp Asp Gly
 20 25 30
 cca tgt aaa gca atc atg aaa aga ttt tct tca ata ttt tca ctc gac 483
 Pro Cys Lys Ala Ile Met Lys Arg Phe Ser Ser Ile Phe Ser Leu Asp
 35 40 45
 agt gcg 489
 Ser Ala
 50

<210> 571
 <211> 484
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 256..483

<220>
 <221> sig_peptide
 <222> 256..327
 <223> Von Heijne matrix
 score 7.09999990463257
 seq FLFLFNLIFFILG/AV

<400> 571
 ggggcggggc ctgccgagtc cgcggcgcttc cccggctgca gccgggaggg ggccgaggag 60
 tgactgagcc ccgggctgtg cagtcgagc ccgactgagg cacgagcggg tgacgctggg 120
 cctgcagcgc ggagcagaaa gcagaaccgc cagagtcctc cctgctgctg tgtggacgac 180
 acgtgggcac aggcagaagt gggccctgtg accagctgca ctggtttcgt ggaaggaagc 240
 tccaggactg gcggg atg ggc tca gcc tgt atc aaa gtc acc aaa tac ttt 291
 Met Gly Ser Ala Cys Ile Lys Val Thr Lys Tyr Phe
 -20 -15
 ctc ttc ctc ttc aac ttg atc ttc ttt atc ctg ggc gca gtg atc ctg 339
 Leu Phe Leu Phe Asn Leu Ile Phe Phe Ile Leu Gly Ala Val Ile Leu
 -10 -5 1

308

```

ggc ttc ggg gtg tgg atc ctg gcc gac aag agc agt ttc atc tct gtc      387
Gly Phe Gly Val Trp Ile Leu Ala Asp Lys Ser Ser Phe Ile Ser Val
5          10          15          20
ctg caa acc tcc tcc agc tcg ctt agg atg ggg gcc tat gtc ttc atc      435
Leu Gln Thr Ser Ser Ser Ser Leu Arg Met Gly Ala Tyr Val Phe Ile
          25          30          35
ggc gtg ggg gca gtc act atg ctc atg ggc ttc ctg ggc tgc atc ggc g      484
Gly Val Gly Ala Val Thr Met Leu Met Gly Phe Leu Gly Cys Ile Gly
          40          45          50

```

<210> 572
 <211> 482
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 29..481

<220>
 <221> sig_peptide
 <222> 29..82
 <223> Von Heijne matrix
 score 7
 seq VLGLVLLSVTVQG/KV

```

<400> 572
agcctagcac tctgacctag cagtcaac atg aag gct ctc att gtt ctg ggg      52
                        Met Lys Ala Leu Ile Val Leu Gly
                        -15
ctt gtc ctc ctt tct gtt acg gtc cag ggc aag gtc ttt gaa agg tgt      100
Leu Val Leu Leu Ser Val Thr Val Gln Gly Lys Val Phe Glu Arg Cys
-10          -5          1          5
gag ttg gcc aga act ctg aaa aga ttg gga atg gat ggc tac agg gga      148
Glu Leu Ala Arg Thr Leu Lys Arg Leu Gly Met Asp Gly Tyr Arg Gly
          10          15          20
atc agc cta gca aac tgg atg tgt ttg gcc aaa tgg gag agt ggt tac      196
Ile Ser Leu Ala Asn Trp Met Cys Leu Ala Lys Trp Glu Ser Gly Tyr
          25          30          35
aac aca cga gct aca aac tac aat gct gga gac aga agc act gat tat      244
Asn Thr Arg Ala Thr Asn Tyr Asn Ala Gly Asp Arg Ser Thr Asp Tyr
          40          45          50
ggg ata ttt cag atc aat agc cgc tac tgg tgt aat gat ggc aaa acc      292
Gly Ile Phe Gln Ile Asn Ser Arg Tyr Trp Cys Asn Asp Gly Lys Thr
          55          60          65          70
cca gga gca gtt aat gcc tgt cat tta tcc tgc agt gct ttg ctg caa      340
Pro Gly Ala Val Asn Ala Cys His Leu Ser Cys Ser Ala Leu Leu Gln
          75          80          85
gat aac atc gct gat gct gta gct tgt gca aag agg gtt gtc cgt gat      388
Asp Asn Ile Ala Asp Ala Val Ala Cys Ala Lys Arg Val Val Arg Asp
          90          95          100
cca ccc acc tcg gcc tcc caa agt gct ggg att aca ggc gtg asc aac      436
Pro Pro Thr Ser Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Xaa Asn
          105          110          115
tgc gcc cgg cca cat tca gtt ctt atc aaa gaa ata acc cag act t      482
Cys Ala Arg Pro His Ser Val Leu Ile Lys Glu Ile Thr Gln Thr
          120          125          130

```

<210> 573
 <211> 541
 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS

<222> 294..539

<220>

<221> sig_peptide

<222> 294..380

<223> Von Heijne matrix

score 7

seq LLMLFHTMAQIMA/EQ

<400> 573

ataactatgt	gacattactt	agggcataat	tctaaaattt	ctgtcctagg	agacactaga	60
taactctcaa	aaacttctga	ttttcttgca	tccctttttg	gcagagtggg	cacagtataa	120
tgcgggaggg	cctcctcccc	cgaatcagca	acaatttttg	ctccaccgtt	ttctttcggg	180
aacagatgcg	ttttcagaag	acagtactcc	agtgataagc	ggcgactttg	agggattccc	240
tctctggcgg	cctctgcagc	agcacagccg	gcctcattcg	gggcaactgcg	agt atg	296
					Met	
gat ctc caa gga aga ggg gtc ccc agc atc gac aga ctt cga gtt ctc						344
Asp Leu Gln Gly Arg Gly Val Pro Ser Ile Asp Arg Leu Arg Val Leu						
	-25		-20		-15	
ctg atg ttg ttc cat aca atg gct caa atc atg gca gaa caa gaa gtg						392
Leu Met Leu Phe His Thr Met Ala Gln Ile Met Ala Glu Gln Glu Val						
	-10		-5		1	
gaa aat ctc tca ggc ctt tcc act aac cct gaa aaa gat ata ttt gtg						440
Glu Asn Leu Ser Gly Leu Ser Thr Asn Pro Glu Lys Asp Ile Phe Val						
5	10		15		20	
gtg cgg gaa aat ggg acg acc tgt ctc atg gca gag ttt gca gcc aaa						488
Val Arg Glu Asn Gly Thr Thr Cys Leu Met Ala Glu Phe Ala Ala Lys						
	25		30		35	
ttt att gta cct tat gat gtg tgg gcc agc aac tac gta gat ctg atc						536
Phe Ile Val Pro Tyr Asp Val Trp Ala Ser Asn Tyr Val Asp Leu Ile						
	40		45		50	
aca ga						541
Thr						

<210> 574

<211> 453

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 274..453

<220>

<221> sig_peptide

<222> 274..345

<223> Von Heijne matrix

score 6.59999990463257

seq KMVHLLVLSGAWG/MQ

<400> 574

attattggtt	gggggaaacc	acgaggggac	gcggccgagg	agggtcgctg	tcaccccggg	60
ggcgtgggag	tgaggyacca	gattcagccc	atttggcccc	gacgcctctg	ttctcggaat	120
ccgggtgctg	cggattgagg	tcccggttcc	taacgaatct	ctgctggatt	ggccgtaacc	180
ctgtccccga	gcgggctcac	agggctctgaa	ggccacgcay	gaggcaaagg	yaaagyycyg	240
agccaccccg	tgchtcsttc	ccaggascgc	aag atg gag gaa ggc ggg aac cta			294
			Met Glu Glu Gly Gly Asn Leu			
			-20			
gga ggc ctg att aag atg gtc cat cta ctg gtc ttg tca ggt gcc tgg						342
Gly Gly Leu Ile Lys Met Val His Leu Leu Val Leu Ser Gly Ala Trp						
	-15		-10		-5	

310

```

ggc atg caa atg tgg gtg acc ttc gtc tca ggc ttc ctg ctt ttc cga      390
Gly Met Gln Met Trp Val Thr Phe Val Ser Gly Phe Leu Leu Phe Arg
  1          5          10          15
agc ctt ccc cga cat acc ttc gga cta gtg cag agc aaa ctc ttc ccc      438
Ser Leu Pro Arg His Thr Phe Gly Leu Val Gln Ser Lys Leu Phe Pro
          20          25          30
ttc tac ttc cac atc      453
Phe Tyr Phe His Ile
          35

```

<210> 575
 <211> 497
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 70..495

<220>
 <221> sig_peptide
 <222> 70..147
 <223> Von Heijne matrix
 score 6.09999990463257
 seq IGLASLFTENVLA/HP

```

<400> 575
atagatgaa accatggaag tcttsctsag atctggggag ggagcctggc ccaagagccc      60
gtggagcta atg aaa aca gaa cag ctg aat aga ttt gct gga ttt ggt att      111
      Met Lys Thr Glu Gln Leu Asn Arg Phe Ala Gly Phe Gly Ile
      -25          -20          -15
gga ctt gca agt ctc ttt aca gaa aat gta ttg gca cat cct tgc att      159
Gly Leu Ala Ser Leu Phe Thr Glu Asn Val Leu Ala His Pro Cys Ile
      -10          -5          1
gtt cta cgc cgc caa tgt cag gtt aat tac cat gct cag cat tac cat      207
Val Leu Arg Arg Gln Cys Gln Val Asn Tyr His Ala Gln His Tyr His
  5          10          15          20
ctc act cca ttt aca gtc atc aat att atg tac agt ttc aac aaa act      255
Leu Thr Pro Phe Thr Val Ile Asn Ile Met Tyr Ser Phe Asn Lys Thr
          25          30          35
cag gga cct aga gcc ctg tgg aaa gga atg rga agt aca ttt att gtc      303
Gln Gly Pro Arg Ala Leu Trp Lys Gly Met Xaa Ser Thr Phe Ile Val
          40          45          50
cag gga gtc aca ctt gga gca gaa ggc ata att agt gaa ttt aca cct      351
Gln Gly Val Thr Leu Gly Ala Glu Gly Ile Ile Ser Glu Phe Thr Pro
          55          60          65
ttg cca agg gag gtt tta cat aaa tgg agt cct aaa caa ata gga gaa      399
Leu Pro Arg Glu Val Leu His Lys Trp Ser Pro Lys Gln Ile Gly Glu
          70          75          80
cac ctt cta ctg aaa tcc cta act tac gtg gtg gca atg ctt ttt tat      447
His Leu Leu Leu Lys Ser Leu Thr Tyr Val Val Ala Met Leu Phe Tyr
          85          90          95          100
tca gca agt ctg att gaa aca gtg cag agt gag ata att cga gat aat      495
Ser Ala Ser Leu Ile Glu Thr Val Gln Ser Glu Ile Ile Arg Asp Asn
          105          110          115
aa

```

<210> 576
 <211> 467
 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS

<222> 88..465

<220>

<221> sig_peptide

<222> 88..219

<223> Von Heijne matrix

score 6.09999990463257

seq ACCLVMVISCVFC/MG

<400> 576

tcttacacca acctgctcca aaccacaaga ggagttactt gttccagcct cctgtgtgga 60
 ctgcttttct atcaaagcac cttagac atg cac gag gaa gaa ata tac acc tct 114
 Met His Glu Glu Glu Ile Tyr Thr Ser

-40

ctt cag tgg gat agc cca gca cca gac act tac cag aaa tgt ctg tct 162
 Leu Gln Trp Asp Ser Pro Ala Pro Asp Thr Tyr Gln Lys Cys Leu Ser

-35 -30 -25 -20
 tcc aac aaa tgt tca gga gca tgc tgt ctt gtg atg gtg att tca tgt 210
 Ser Asn Lys Cys Ser Gly Ala Cys Cys Leu Val Met Val Ile Ser Cys

-15 -10 -5
 gtt ttc tgc atg gga tta tta acg gca tcc att ttc ttg ggc gtc aag 258
 Val Phe Cys Met Gly Leu Leu Thr Ala Ser Ile Phe Leu Gly Val Lys

1 5 10
 ttg ttg cag gtg tcc acc att gcg atg cag cag caa gaa aaa ctc atc 306
 Leu Leu Gln Val Ser Thr Ile Ala Met Gln Gln Gln Glu Lys Leu Ile

15 20 25
 caa caa gag agg gca ctg cta aac ttt aca gaa tgg aag aga agc tgt 354
 Gln Gln Glu Arg Ala Leu Leu Asn Phe Thr Glu Trp Lys Arg Ser Cys

30 35 40 45
 gcc ctt cag atg aaa tat tgc caa gcc ttc atg caa aac tca tta agt 402
 Ala Leu Gln Met Lys Tyr Cys Gln Ala Phe Met Gln Asn Ser Leu Ser

50 55 60
 tca gcc cat aac agc agt cct tgt cca aac aat tgg att cag aac aga 450
 Ser Ala His Asn Ser Ser Pro Cys Pro Asn Asn Trp Ile Gln Asn Arg

65 70 75
 gaa agt tgt tac tat gt 467
 Glu Ser Cys Tyr Tyr

80

<210> 577

<211> 485

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 328..483

<220>

<221> sig_peptide

<222> 328..387

<223> Von Heijne matrix

score 5.90000009536743

seq GWMAAGLMIGAGA/CY

<400> 577

ctagagcaca agtcctgcta gtcgcctccg tctgggtacc agccccctat tactctgcag 60
 gcgtgtgaag aaagaaggaa actagctcgg accgtgcagg tttgtaggtc tgttggcctg 120
 taggtttcgg cacaagtttc agcgagagaa ggagaaaact gccttggttg gaaccttgca 180
 gattcatcac aagagagcta caagagcctg gaagaagctg aagactgcta cctccatcc 240
 ttactcacc tggacctgag agacctcttc aatcaggttg agcaaggccc tctcctgtcc 300
 tgccccaagg ctggcacaga cttgagc atg ggc cgg gct cgg gaa gtg ggt tgg 354

312

								Met	Gly	Arg	Ala	Arg	Glu	Val	Gly	Trp	
								-20							-15		
atg	gcg	gca	gga	ctg	atg	att	ggg	gct	ggg	gcc	tgc	tac	tgc	gtt	tac		402
Met	Ala	Ala	Gly	Leu	Met	Ile	Gly	Ala	Gly	Ala	Cys	Tyr	Cys	Val	Tyr		
	-10					-5					1				5		
aaa	ctg	acc	ata	gga	aga	gat	gac	agt	gag	aag	ctg	gag	gag	gag	ggg		450
Lys	Leu	Thr	Ile	Gly	Arg	Asp	Asp	Ser	Glu	Lys	Leu	Glu	Glu	Glu	Gly		
			10						15					20			
gaa	gag	gag	tgg	gac	gat	gac	cag	gag	ctg	gat	ga						485
Glu	Glu	Glu	Trp	Asp	Asp	Asp	Gln	Glu	Leu	Asp							
			25					30									

<210> 578

<211> 502

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 39..500

<220>

<221> sig_peptide

<222> 39..83

<223> Von Heijne matrix

score 5.90000009536743

seq ILMLTFIICGLLT/RV

<400> 578

attcctcagg	acacagagct	tcctctctcc	caggagcc	atg	aat	atc	ctg	atg	ctg								56
				Met	Asn	Ile	Leu	Met	Leu								
				-15					-10								
acc	ttc	att	atc	tgt	ggg	ttg	cta	act	cgg	gtg	acc	aaa	ggg	agc	ttt		104
Thr	Phe	Ile	Ile	Cys	Gly	Leu	Leu	Thr	Arg	Val	Thr	Lys	Gly	Ser	Phe		
				-5					1				5				
gaa	ccc	caa	aaa	tgt	tgg	aag	aat	aat	gta	gga	cat	tgc	aga	aga	cga		152
Glu	Pro	Gln	Lys	Cys	Trp	Lys	Asn	Asn	Val	Gly	His	Cys	Arg	Arg	Arg		
		10					15					20					
tgt	tta	gat	act	gaa	agg	tac	ata	ctt	ctt	tgt	agg	aac	aag	cta	tca		200
Cys	Leu	Asp	Thr	Glu	Arg	Tyr	Ile	Leu	Leu	Cys	Arg	Asn	Lys	Leu	Ser		
		25				30					35						
tgc	tgc	att	tct	ata	ata	tca	cat	gaa	tat	act	cga	cga	cca	gca	ttt		248
Cys	Cys	Ile	Ser	Ile	Ile	Ser	His	Glu	Tyr	Thr	Arg	Arg	Pro	Ala	Phe		
				45					50						55		
cct	gtg	att	cac	cta	gag	gat	ata	aca	ttg	gat	tat	agt	gat	gtg	gac		296
Pro	Val	Ile	His	Leu	Glu	Asp	Ile	Thr	Leu	Asp	Tyr	Ser	Asp	Val	Asp		
			60						65					70			
tct	ttt	act	ggg	tcc	cca	gta	tct	atg	ttg	aat	gat	ctg	ata	aca	ttt		344
Ser	Phe	Thr	Gly	Ser	Pro	Val	Ser	Met	Leu	Asn	Asp	Leu	Ile	Thr	Phe		
			75					80					85				
gac	aca	act	aaa	ttt	gga	gaa	acc	atg	aca	cct	gag	acc	aat	act	cct		392
Asp	Thr	Thr	Lys	Phe	Gly	Glu	Thr	Met	Thr	Pro	Glu	Thr	Asn	Thr	Pro		
			90					95					100				
gag	act	act	atg	cca	cca	tcc	gag	gcc	act	act	ccc	gag	act	act	atg		440
Glu	Thr	Thr	Met	Pro	Pro	Ser	Glu	Ala	Thr	Thr	Pro	Glu	Thr	Thr	Met		
		105				110					115						
cca	cca	tct	gag	act	gct	act	tcc	gag	act	atg	cca	cca	cct	tct	cag		488
Pro	Pro	Ser	Glu	Thr	Ala	Thr	Ser	Glu	Thr	Met	Pro	Pro	Pro	Ser	Gln		
					125					130					135		
aca	gct	ctt	act	ca													502
Thr	Ala	Leu	Thr														

<210> 579

<211> 754
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 259..753

<220>
 <221> sig_peptide
 <222> 259..375
 <223> Von Heijne matrix
 score 5.80000019073486
 seq FCVCVIAIGVVQA/LI

<400> 579
 aagctcccg cgggctgac tcaagcggag ggcgcgggaa cagtcgccga ggcgattccc 60
 gccagcagt tcgacagaag tgtacagagg cttctggcaa cacggattgc cgtctacctg 120
 atgacctttc tcacgtgac agtggcctgg gcagcacaca caaggttgtt ccaagttgtt 180
 gggaaaacag acgacacact tgccctgctc aacctggccg catcatggct gtgatgccct 240
 ccctccctcc aggctgac atg atg acc atc acc ttc ctg cct tac acg ttt 291
 Met Met Thr Ile Thr Phe Leu Pro Tyr Thr Phe
 -35 -30
 tcg tta atg gtg acc ttc cct gat gtg cct ctg ggc atc ttc ttg ttc 339
 Ser Leu Met Val Thr Phe Pro Asp Val Pro Leu Gly Ile Phe Leu Phe
 -25 -20 -15
 tgt gtg tgt gtg atc gcc atc ggg gtc gtg cag gca ctg att gtr ggg 387
 Cys Val Cys Val Ile Ala Ile Gly Val Val Gln Ala Leu Ile Val Gly
 -10 -5 1
 tac gca ttc cac ttc ccg cac ctg ctg agc ccg cag atc cag cgc tct 435
 Tyr Ala Phe His Phe Pro His Leu Leu Ser Pro Gln Ile Gln Arg Ser
 5 10 15 20
 gcc cac agg gct ctg tac cga cga cac gtc ctg ggc atc gtc ctc caa 483
 Ala His Arg Ala Leu Tyr Arg Arg His Val Leu Gly Ile Val Leu Gln
 25 30 35
 ggc ccg gcc ctg tgc ttt gca gcg gcc atc ttc tct ctc ttc ttt gtc 531
 Gly Pro Ala Leu Cys Phe Ala Ala Ala Ile Phe Ser Leu Phe Phe Val
 40 45 50
 ccc ttg tct tac ctg ctg atg gtg act gtc atc ctc ctc ccc tat gtc 579
 Pro Leu Ser Tyr Leu Leu Met Val Thr Val Ile Leu Leu Pro Tyr Val
 55 60 65
 agc aag gtc acc ggc tgg tgc aga gac agg ctc ctg ggc cac agg gag 627
 Ser Lys Val Thr Gly Trp Cys Arg Asp Arg Leu Leu Gly His Arg Glu
 70 75 80
 ccc tcg gct cac cca gtg gaa gtc ttc tcg ttt gac ctc cac gag cca 675
 Pro Ser Ala His Pro Val Glu Val Phe Ser Phe Asp Leu His Glu Pro
 85 90 95 100
 ctc agc aag gag cgc gtg gaa gcc ttc agc gac gga gtc tac gcc atc 723
 Leu Ser Lys Glu Arg Val Glu Ala Phe Ser Asp Gly Val Tyr Ala Ile
 105 110 115
 gtg gcc acg ctt ctc atc ctg gac atc tgg t 754
 Val Ala Thr Leu Leu Ile Leu Asp Ile Trp
 120 125

<210> 580
 <211> 488
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 54..488

<220>

<221> sig_peptide

<222> 54..119

<223> Von Heijne matrix

score 5.80000019073486

seq PVTLTSLLPMTX/GP

<220>

<221> misc_feature

<222> 181,256,369

<223> n=a, g, c or t

<400> 580

```

gtcttttttaa tctctgcagt cccctgagtg tcacccaacc tgcaggcagt gtc atg      56
                                     Met
ggc cgt tgg agt ctg act gca tct cct gtt acc ctc aca tct ctc tta      104
Gly Arg Trp Ser Leu Thr Ala Ser Pro Val Thr Leu Thr Ser Leu Leu
-20                               -15          -10
cca atg gta act kca gga cca gct gca ggg aag agc agt tcc tca acc      152
Pro Met Val Thr Xaa Gly Pro Ala Ala Gly Lys Ser Ser Ser Ser Thr
-5                               1              5              10
tcg tgg gat act gtg ctg act gcc atc wnc tgt gcc agc act gtg cag      200
Ser Trp Asp Thr Val Leu Thr Ala Ile Xaa Cys Ala Ser Thr Val Gln
15                               20              25
ctg atc tcc aca aca ctg gga gca tct gcc tca ggt gcc aga atg ccc      248
Leu Ile Ser Thr Thr Leu Gly Ala Ser Ala Ser Gly Ala Arg Met Pro
30                               35              40
act acc tnc tgc tcg ggg acc act gtg ttc ctg act gcc ctt cag gat      296
Thr Thr Xaa Cys Ser Gly Thr Thr Val Phe Leu Thr Ala Leu Gln Asp
45                               50              55
act atg cag aga gag gag ctt gta aaa aat gcc act cct cct gca gaa      344
Thr Met Gln Arg Glu Glu Leu Val Lys Asn Ala Thr Pro Pro Ala Glu
60                               65              70              75
cct gcc agg gca gag gac ctt tct nct gct cct cat gtg aca cca acc      392
Pro Ala Arg Ala Glu Asp Leu Ser Xaa Ala Pro His Val Thr Pro Thr
80                               85              90
tcg tgc tgt ccc aca ctg gca cct gca gca cca cct gct tcc ctg ggc      440
Ser Cys Cys Pro Thr Leu Ala Pro Ala Ala Pro Pro Ala Ser Leu Gly
95                               100              105
act atc ttg atg aca atc atg ttt gcc agc gta ggt ttt tcc aat gaa      488
Thr Ile Leu Met Thr Ile Met Phe Ala Ser Val Gly Phe Ser Asn Glu
110                               115              120

```

<210> 581

<211> 343

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 17..343

<220>

<221> sig_peptide

<222> 17..106

<223> Von Heijne matrix

score 5.69999980926514

seq LGVNLVVMPPTRA/RS

<400> 581

```

atagattaca acggtg atg gcg ggc aag cgg tcc ggc tgg agc cgg gcg gct      52
Met Ala Gly Lys Arg Ser Gly Trp Ser Arg Ala Ala
-30                               -25              -20

```

315

ctc	ctc	cag	ctc	ctt	ctc	ggc	gtg	aac	ctg	gtg	gtg	atg	ccg	ccc	acc	100
Leu	Leu	Gln	Leu	Leu	Leu	Gly	Val	Asn	Leu	Val	Val	Met	Pro	Pro	Thr	
		-15						-10					-5			
cgg	gcc	cgg	agt	ctg	cgc	ttc	gtt	acc	ttg	gag	ggg	atg	cta	cag	cac	148
Arg	Ala	Arg	Ser	Leu	Arg	Phe	Val	Thr	Leu	Glu	Gly	Met	Leu	Gln	His	
	1					5					10					
tgg	gaa	ctg	ggc	cag	gcc	ctg	cgg	cag	cgc	tat	cac	ggc	ttc	cta	aac	196
Trp	Glu	Leu	Gly	Gln	Ala	Leu	Arg	Gln	Arg	Tyr	His	Gly	Phe	Leu	Asn	
15				20					25					30		
acc	tct	tat	cac	cgg	caa	gag	gtt	tat	gtg	cga	arg	tac	ttt	gac	cgg	244
Thr	Ser	Tyr	His	Arg	Gln	Glu	Val	Tyr	Val	Arg	Xaa	Tyr	Phe	Asp	Arg	
			35					40						45		
act	ctc	atg	agt	gct	gag	gcc	aac	ctg	gct	gga	stc	ttc	cct	ccc	aac	292
Thr	Leu	Met	Ser	Ala	Glu	Ala	Asn	Leu	Ala	Gly	Xaa	Phe	Pro	Pro	Asn	
		50					55					60				
ggg	atk	cag	cgc	ttc	aac	ccg	aac	atc	tcg	tgg	cag	cct	att	cct	gtg	340
Gly	Xaa	Gln	Arg	Phe	Asn	Pro	Asn	Ile	Ser	Trp	Gln	Pro	Ile	Pro	Val	
	65					70						75				
cac																343
His																

<210> 582

<211> 465

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 293..463

<220>

<221> sig_peptide

<222> 293..388

<223> Von Heijne matrix

score 5.59999990463257

seq WMGLACFRSLAAS/SP

<400> 582

aaaagattga	agaggaactc	tttcaaacct	ttcaagtgcc	ccctcctttc	cttaaagtct	60
tttatagggg	tccccttctt	ggccatctcc	atcctgtgag	tcaggactga	aagggcacag	120
acaggtcact	gccagcattg	ttggggcaag	cctgcaagca	cgcactcactg	gggatctgac	180
atgacaatgg	ccgcctgccc	cctctgaggg	ctacaggact	tacccagtg	ggaagcagct	240
aagcaggtct	gaccagccga	cctggacctg	gccaagggtc	ctgtcatccc	tc atg gcc	298
					Met Ala	
acc ccg cca ttc cgg ctg ata agg aag atg ttt tcc ttc aag gtg agc						346
Thr Pro Pro Phe Arg Leu Ile Arg Lys Met Phe Ser Phe Lys Val Ser						
-30		-25		-20		-15
aga tgg atg ggg ctt gcc tgc ttc cgg tcc ctg gcg gca tcc tct ccc						394
Arg Trp Met Gly Leu Ala Cys Phe Arg Ser Leu Ala Ala Ser Ser Pro						
	-10		-5		1	
agt att cgc cag aag aaa cta atg cac aag ctg cag gag gaa aac gct						442
Ser Ile Arg Gln Lys Lys Leu Met His Lys Leu Gln Glu Glu Asn Ala						
	5		10		15	
ttt cgc gaa gag atg aaa att tt						465
Phe Arg Glu Glu Met Lys Ile						
20		25				

<210> 583

<211> 539

<212> DNA

<213> Homo sapiens

<220>

<221> CDS
<222> 302..538

<220>

<221> sig_peptide

<222> 302..358

<223> Von Heijne matrix

score 5.40000009536743

seq LCCSGCVPSLCCS/SY

<400> 583

```

aaaaacttcc gccgccgcgt ccgccgcctc cggaactaaa cgggggtgagg tcacattcgg      60
ttatctctaa cgttggaaaa cgatggagct aacacccatt atggagatta accacttttc      120
atcaggtttt taacttaagt cgtgaggaat acaacggtga acacaagatt cattttattt      180
tcatcaccat gggacgtatc ctggttgtga gtwctctggg tcagacctct gaagacttct      240
cagatggatc ctagtctctg ggcttgccctg aaattactcg ctgctcaggg agagagttga      300
a atg gtt ggc atc ctc cca ctc tgt tgc tcc ggc tgt gtc ccc tcg ctc      349
  Met Val Gly Ile Leu Pro Leu Cys Cys Ser Gly Cys Val Pro Ser Leu
        -15                    -10                    -5

tgt tgt tcc agc tat gtc ccc tct gtt gct cca act gca gct cat tct      397
Cys Cys Ser Ser Tyr Val Pro Ser Val Ala Pro Thr Ala Ala His Ser
        1                    5                    10

gtt aga gtt cct cat tca gct ggt cac tgt ggc cag agg gtg ttg gcc      445
Val Arg Val Pro His Ser Ala Gly His Cys Gly Gln Arg Val Leu Ala
        15                    20                    25

tgc tcc ctt cct caa gta ttc tta aag cca tgg att ttt gtg gag cat      493
Cys Ser Leu Pro Gln Val Phe Leu Lys Pro Trp Ile Phe Val Glu His
        30                    35                    40                    45

ttt tct tcc tgg ctc tcc ctt gag tta ttt tcc ttt ctt cgc tat c      539
Phe Ser Ser Trp Leu Ser Leu Glu Leu Phe Ser Phe Leu Arg Tyr
        50                    55                    60

```

<210> 584

<211> 526

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 85..525

<220>

<221> sig_peptide

<222> 85..144

<223> Von Heijne matrix

score 5.19999980926514

seq SPVFLVFPPEITA/SE

<400> 584

```

cctartcgaa aaacaagaaa agaacatggt ctagactgaa gtaccaacta aatcatctcc      60
tttcaaatta tcaccgacac catc atg gat tca agc acc gca cac agt ccg      111
        Met Asp Ser Ser Thr Ala His Ser Pro
        -20                    -15

gtg ttt ctg gta ttt cct cca gaa atc act gct tca gaa tat gag tcc      159
Val Phe Leu Val Phe Pro Pro Glu Ile Thr Ala Ser Glu Tyr Glu Ser
        -10                    -5                    1                    5

aca gaa ctt tca gcc acg acc ttt tca act caa agc ccc ttg caa aaa      207
Thr Glu Leu Ser Ala Thr Thr Phe Ser Thr Gln Ser Pro Leu Gln Lys
        10                    15                    20

tta ttt gct aga aaa atg aaa atc tta ggg act atc cag atc ctg ttt      255
Leu Phe Ala Arg Lys Met Lys Ile Leu Gly Thr Ile Gln Ile Leu Phe
        25                    30                    35

gga att atg acc ttt tct ttt gga gtt atc ttc ctt ttc act ttg tta      303

```

317

Gly	Ile	Met	Thr	Phe	Ser	Phe	Gly	Val	Ile	Phe	Leu	Phe	Thr	Leu	Leu	
		40					45					50				
aaa	cca	tat	cca	agg	ttt	ccc	ttt	ata	ttt	ctt	tca	gga	tat	cca	ttc	351
Lys	Pro	Tyr	Pro	Arg	Phe	Pro	Phe	Ile	Phe	Leu	Ser	Gly	Tyr	Pro	Phe	
		55					60					65				
tgg	ggc	tct	gtt	ttg	ttc	att	aat	tct	gga	gcc	ttc	cta	att	gca	gtg	399
Trp	Gly	Ser	Val	Leu	Phe	Ile	Asn	Ser	Gly	Ala	Phe	Leu	Ile	Ala	Val	
					75					80					85	
aaa	aga	aaa	acc	aca	gaa	act	ctg	ata	ata	ttg	agc	cga	ata	atg	aat	447
Lys	Arg	Lys	Thr	Glu	Thr	Leu	Ile	Ile	Leu	Ser	Arg	Ile	Met	Asn		
					90				95					100		
ttt	ctt	agt	gcc	ctg	gga	gca	ata	gct	gga	atc	att	ctc	ctc	aca	ttt	495
Phe	Leu	Ser	Ala	Leu	Gly	Ala	Ile	Ala	Gly	Ile	Ile	Leu	Leu	Thr	Phe	
			105					110						115		
ggg	ttt	cat	cct	aga	tca	aaa	cta	cat	ttg	t						526
Gly	Phe	His	Pro	Arg	Ser	Lys	Leu	His	Leu							
		120					125									

<210> 585

<211> 482

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 195..482

<220>

<221> sig_peptide

<222> 195..293

<223> Von Heijne matrix

score 5.09999990463257

seq LYALVQLGQPCDC/LP

<400> 585

gcacagatga	ccggaagcgt	gccctggcgg	cgctgccagc	cccagctggg	agctggtagg		60
ggagtgcgag	gaaggcggga	cgcaggcgcg	agccgctgca	ggggcggggc	ctgcggacag		120
gagccggggg	ttggggcggg	aacccctcgt	cccctgcaga	ccccgcctgc	tcgggcgcgg		180
gcggcggcgc	ggcc atg aag	ctg aag	ctg aag aac	gtg ttt ctc	gcc tac		230
	Met Lys Leu	Lys Leu	Lys Asn Val	Phe Leu	Ala Tyr		
		-30		-25			
ttc ctg gtg	tgc atc gcc	ggc ctc	ctc tac	gcg ctg	gta cag	ctc ggc	278
Phe Leu Val	Ser Ile Ala	Gly Leu	Leu Tyr	Ala Leu	Val Gln	Leu Gly	
	-20	-15		-10			
cag cca tgt	gac tgc	ctt cct	ccc ctg	cgg gca	gcs gag	cag cta	326
Gln Pro Cys	Asp Cys	Leu Pro	Pro Leu	Arg Ala	Ala Glu	Gln Leu	
	-5	1		5		10	
cgg cag aag	gat ctg	agg att	tcc cag	ctg caa	gcg gaa	ctc cga	374
Arg Gln Lys	Asp Leu	Arg Ile	Ser Gln	Leu Gln	Ala Glu	Leu Arg	
	15		20		25		
cca ccc cct	gcc cct	gcc cag	ccc cct	gaa ccc	gag gcc	ctg cct	422
Pro Pro Pro	Ala Pro	Ala Gln	Pro Pro	Glu Pro	Glu Ala	Leu Pro	
	30		35		40		
atc tat gtt	gtt acc	ccc acc	tat gcc	agg ctg	gta cag	aag gca	470
Ile Tyr Val	Val Thr	Pro Thr	Tyr Ala	Arg Leu	Val Gln	Lys Ala	
	45		50		55		
ctg gta cga	ctg						482
Leu Val Arg	Leu						
	60						

<210> 586

<211> 543

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 35..541

<220>

<221> sig_peptide

<222> 35..82

<223> Von Heijne matrix

score 5

seq CGLFLRTTAAARA/CR

<400> 586

```

agactgaggc tgaggctggg gaacatcggg cagc atg agc ggc tgc ggg ctc ttc      55
                                Met Ser Gly Cys Gly Leu Phe
                                -15                                -10
ctg cgc acc acg gct gcg gct cgt gcc tgc cgg ggt ctg gtg gtc tct      103
Leu Arg Thr Thr Ala Ala Arg Ala Cys Arg Gly Leu Val Val Ser
                                -5                                1                                5
acc gcg aac cgg cgg cta ctg cgc acc agc ccg cct gta cga gct ttc      151
Thr Ala Asn Arg Arg Leu Leu Arg Thr Ser Pro Pro Val Arg Ala Phe
                                10                                15                                20
gcc aaa gag ctt ttc cta ggc aaa atc aag aag aaa gaa gtt ttc cca      199
Ala Lys Glu Leu Phe Leu Gly Lys Ile Lys Lys Lys Glu Val Phe Pro
                                25                                30                                35
ttt cca gaa gtt agc caa gat gaa ctt aat gaa atc aat cag ttc ttg      247
Phe Pro Glu Val Ser Gln Asp Glu Leu Asn Glu Ile Asn Gln Phe Leu
40                                45                                50                                55
gga ccc gtg gaa aaa ttc ttc act gaa gag gtg gac tcc cga aaa att      295
Gly Pro Val Glu Lys Phe Phe Thr Glu Glu Val Asp Ser Arg Lys Ile
                                60                                65                                70
gac cag gaa ggg aaa atc cca gat gaa act ttg gag aaa ttg aag agc      343
Asp Gln Glu Gly Lys Ile Pro Asp Glu Thr Leu Glu Lys Leu Lys Ser
                                75                                80                                85
cta ggg ctt ttt ggg ctg caa gtc cca gaa gaa tat ggt ggc ctg ggc      391
Leu Gly Leu Phe Gly Leu Gln Val Pro Glu Glu Tyr Gly Gly Leu Gly
90                                95                                100
ttc tcc aac acc atg tac tca cga cta ggg gag atc atc agc atg gat      439
Phe Ser Asn Thr Met Tyr Ser Arg Leu Gly Glu Ile Ile Ser Met Asp
105                                110                                115
ggg tcc atc act gtg acc ctg gca gcg cac cag gct att ggc tca agg      487
Gly Ser Ile Thr Val Thr Leu Ala Ala His Gln Ala Ile Gly Ser Arg
120                                125                                130                                135
gga tca tct tgg ctg gca ctg agg agc aga aag cca aat aac ttg cct      535
Gly Ser Ser Trp Leu Ala Leu Arg Ser Arg Lys Pro Asn Asn Leu Pro
                                140                                145                                150
aaa ctg gc
Lys Leu

```

<210> 587

<211> 473

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 64..471

<220>

<221> sig_peptide

<222> 64..171

<223> Von Heijne matrix

score 4.80000019073486

seq LLLRLNDAALRA/LQ

<400> 587

```

aggaggtggc ggtggtggcc ctcgcctgtg gcccccggtgc tgcttgcaact cgaactcgtc      60
gcc atg gag gag ctc cag gag cct ctg aga gga gag ctc cgg ctc tgc      108
    Met Glu Glu Leu Gln Glu Pro Leu Arg Gly Glu Leu Arg Leu Cys
      -35                -30                -25
ttc acg caa gct gcc cgg act agc ctc tta ctg ctc agg ctc aac gac      156
Phe Thr Gln Ala Ala Arg Thr Ser Leu Leu Leu Leu Arg Leu Asn Asp
      -20                -15                -10
gct gcc ctg cgg gcg ctg caa gag tgt cag cgg caa cag gta cgg ccg      204
Ala Ala Leu Arg Ala Leu Gln Glu Cys Gln Arg Gln Gln Val Arg Pro
      -5                1                5                10
gtg att gct ttc caa ggc cac cga ggg tat ctg aga ctc cca ggc cct      252
Val Ile Ala Phe Gln Gly His Arg Gly Tyr Leu Arg Leu Pro Gly Pro
      15                20                25
ggg tgg tcc tgc ctc ttc tcc ttc ata gtg tcc cag tgt tgt cag gag      300
Gly Trp Ser Cys Leu Phe Ser Phe Ile Val Ser Gln Cys Cys Gln Glu
      30                35                40
ggc gct ggt ggt agc ttg gac ctt gtg tgc caa cgc ttc ctc agg tct      348
Gly Ala Gly Gly Ser Leu Asp Leu Val Cys Gln Arg Phe Leu Arg Ser
      45                50                55
ggg cct aac agc ctc cac tgc ctg ggc tca ctc agg gag cgc ctc att      396
Gly Pro Asn Ser Leu His Cys Leu Gly Ser Leu Arg Glu Arg Leu Ile
      60                65                70                75
att tgg gca gcc atg gat tct atc cca gcc cca tca tca gtt cag gga      444
Ile Trp Ala Ala Met Asp Ser Ile Pro Ala Pro Ser Ser Val Gln Gly
      80                85                90
cac aay ctg act gaa gat gcc aga cat cc      473
His Asn Leu Thr Glu Asp Ala Arg His
      95                100

```

<210> 588

<211> 509

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 183..509

<220>

<221> sig_peptide

<222> 183..251

<223> Von Heijne matrix

score 4.59999990463257

seq GHSLVGLALLCIA/AN

<400> 588

```

aggaccaatg aaagcaaagt acctcatccc ctcagtgact aagaatcgca gtattttaaga      60
ggtagcagga atgggctgag agtgggtgtt gctttctcca ccagaagggc acactttcat      120
ctaatttggg gtatcactga gctgaagaca aagagaaggg ggagaaaacc tagcagacca      180
cc atg tgc tat ggg aag tgt gca cga tgc atc gga cat tct ctg gtg      227
    Met Cys Tyr Gly Lys Cys Ala Arg Cys Ile Gly His Ser Leu Val
      -20                -15                -10
ggg ctc gcc ctc ctg tgc atc gcg gct aat att ttg ctt tac ttt ccc      275
Gly Leu Ala Leu Leu Cys Ile Ala Ala Asn Ile Leu Leu Tyr Phe Pro
      -5                1                5
aat ggg gaa aca aag tat gcc tcc gaa aac cac ctc agc cgc ttc gtg      323
Asn Gly Glu Thr Lys Tyr Ala Ser Glu Asn His Leu Ser Arg Phe Val
      10                15                20
tgg ttc ttt tct ggc atc gta gga ggt ggc ctg ctg atg ctc ctg cca      371

```


320

Trp	Phe	Phe	Ser	Gly	Ile	Val	Gly	Gly	Gly	Leu	Leu	Met	Leu	Leu	Pro	
25					30					35					40	
gca	ttt	gtc	ttc	att	ggg	ctg	gaa	cag	gat	gac	tgc	tgt	ggc	tgc	tgt	419
Ala	Phe	Val	Phe	Ile	Gly	Leu	Glu	Gln	Asp	Asp	Cys	Cys	Gly	Cys	Cys	
				45					50					55		
ggc	cat	gaa	aac	tgt	ggc	aaa	cga	tgt	gcg	atg	ctt	tct	tct	gta	ttg	467
Gly	His	Glu	Asn	Cys	Gly	Lys	Arg	Cys	Ala	Met	Leu	Ser	Ser	Val	Leu	
			60					65					70			
gct	gct	ctc	att	gga	att	gca	gga	ycr	ggc	tac	tgt	gtc	att			509
Ala	Ala	Leu	Ile	Gly	Ile	Ala	Gly	Xaa	Gly	Tyr	Cys	Val	Ile			
		75					80					85				

<210> 589
 <211> 523
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 302..523

<220>
 <221> sig_peptide
 <222> 302..352
 <223> Von Heijne matrix
 score 4.59999990463257
 seq LLPLCQRIHCCIIG/RF

<400> 589	
agatgagtgt tcagctctca gcagagaggt tagctcctct ctgcagcttg tcctgttgct	60
tcctcaagtc tggctgagtc cggagttttt atgagcctca gaggggagga agtgcattgt	120
gattaatcca tgggcaggcc tggaaaagtt cccactccag tctgcgggac ccacagcctg	180
gccctcaggc tcaggcttcc cwggcttgaa gattgggctt cacctgggac ctacccttc	240
tgcctaggag catgtctgcc tcctgctgcc ttcatgggtg cccagccaag gctaagtcga	300
a atg tgg ccc aac ctt ctt ccc ctg tgc cag cgg atc cat tgc atc att	349
Met Trp Pro Asn Leu Leu Pro Leu Cys Gln Arg Ile His Cys Ile Ile	
-15 -10 -5	
ggt cgc ttc cgg tgc aat ggg ttt gag gac tgt ccc gat ggc agc gat	397
Gly Arg Phe Arg Cys Asn Gly Phe Glu Asp Cys Pro Asp Gly Ser Asp	
1 5 10 15	
gaa gag aac tgc aca gca aac cct ctg ctt tgc tcc acc gcc cgc tac	445
Glu Glu Asn Cys Thr Ala Asn Pro Leu Leu Cys Ser Thr Ala Arg Tyr	
20 25 30	
cac tgc aag aac ggc ctc tgt att gac aag agc ttc atc tgc gat gga	493
His Cys Lys Asn Gly Leu Cys Ile Asp Lys Ser Phe Ile Cys Asp Gly	
35 40 45	
cag aat aac tgt caa gac aac agt gat gag	523
Gln Asn Asn Cys Gln Asp Asn Ser Asp Glu	
50 55	

<210> 590
 <211> 513
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 103..513

<220>
 <221> sig_peptide
 <222> 103..243
 <223> Von Heijne matrix

score 4.40000009536743
seq MATMLQLVPLLKA/SI

<400> 590

```

aatcattaca gaaggatttg gcctctaagg gcctctggaa ttctttggaa gaaggcattg      60
gctgtgtagc tgacactcta ctcactgtgg acatggacgt cc atg acc ctg cag      114
                                   Met Thr Leu Gln
                                   -45
tgg cac ttg tat gag atg gca cgc aac ctg aag gtg cag gat atg ctg      162
Trp His Leu Tyr Glu Met Ala Arg Asn Leu Lys Val Gln Asp Met Leu
                                   -40
                                   -35
                                   -30
cgg gca gag gtc ttg gct gcg cgg cac cag gcc cag gga gac atg gcc      210
Arg Ala Glu Val Leu Ala Ala Arg His Gln Ala Gln Gly Asp Met Ala
                                   -25
                                   -20
                                   -15
acg atg cta cag ctg gtc ccc ctc ctc aaa gcc agc atc aag gag aca      258
Thr Met Leu Gln Leu Val Pro Leu Leu Lys Ala Ser Ile Lys Glu Thr
                                   -10
                                   -5
                                   1
                                   5
cta aga ctt cac ccc atc tcc gtg acc ctg cag aga tat ctt gta aat      306
Leu Arg Leu His Pro Ile Ser Val Thr Leu Gln Arg Tyr Leu Val Asn
                                   10
                                   15
                                   20
gac ttg gtt ctt cga gat tac atg att cct gcc aag aca ctg gtg caa      354
Asp Leu Val Leu Arg Asp Tyr Met Ile Pro Ala Lys Thr Leu Val Gln
                                   25
                                   30
                                   35
gtg gcc atc tat gct ctg ggc cga gag ccc acc ttc ttc ttc gac ccg      402
Val Ala Ile Tyr Ala Leu Gly Arg Glu Pro Thr Phe Phe Phe Asp Pro
                                   40
                                   45
                                   50
gaa aat ttt gac cca acc cga tgg ctg agc aaa gac aag aac atc acc      450
Glu Asn Phe Asp Pro Thr Arg Trp Leu Ser Lys Asp Lys Asn Ile Thr
                                   55
                                   60
                                   65
tac ttc cgg aac ttg ggc ttt ggc tgg ggt gtg cgg cag tgt ctg gga      498
Tyr Phe Arg Asn Leu Gly Phe Gly Trp Gly Val Arg Gln Cys Leu Gly
                                   70
                                   75
                                   80
                                   85
cgg csg atc gct gag      513
Arg Xaa Ile Ala Glu
                                   90

```

<210> 591

<211> 490

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 196..489

<220>

<221> sig_peptide

<222> 196..255

<223> Von Heijne matrix

score 4.30000019073486

seq LLDSSLMASGTAS/RS

<400> 591

```

agacgtacct cacggaagcc ggctttggcc ctgcggctgc taccgtcgcc gcggagaaat      60
tggttgatct ggcagtctag gaatgaatct cctctcagcc ttaagctca cctggtcaga      120
atccttgat gagcctgtgg gaccgttcct cctagcccg tggtttggaa ccagtggctt      180
tgggactgta agagg atg gac aaa gat tct cag ggg ctg cta gat tca tcc      231
                                   Met Asp Lys Asp Ser Gln Gly Leu Leu Asp Ser Ser
                                   -20
                                   -15
                                   -10
ctg atg gca tca ggc act gcc agc cgc tca gag gat gag gag tca ctg      279
Leu Met Ala Ser Gly Thr Ala Ser Arg Ser Glu Asp Glu Glu Ser Leu
                                   -5
                                   1
                                   5
gca ggg cag aag cga gcc tcc tcc cag gcc ttg ggc acc atc cct aaa      327

```

322

Ala	Gly	Gln	Lys	Arg	Ala	Ser	Ser	Gln	Ala	Leu	Gly	Thr	Ile	Pro	Lys	
10						15					20					
cgg	aga	agc	tcc	tcc	agg	ttc	atc	aag	agg	aag	aag	ttc	gat	gat	gag	375
Arg	Arg	Ser	Ser	Ser	Arg	Phe	Ile	Lys	Arg	Lys	Lys	Phe	Asp	Asp	Glu	
25					30				35						40	
ctg	gtg	gag	agc	agc	ctg	gca	aaa	tct	tct	acc	cgg	gca	aag	ggg	gcc	423
Leu	Val	Glu	Ser	Ser	Leu	Ala	Lys	Ser	Ser	Thr	Arg	Ala	Lys	Gly	Ala	
				45				50						55		
agt	ggg	gtg	gaa	cag	ggc	gct	gtt	cgg	gta	gtg	aac	cct	cct	cca	gtg	471
Ser	Gly	Val	Glu	Gln	Gly	Ala	Val	Arg	Val	Val	Asn	Pro	Pro	Pro	Val	
			60					65					70			
aga	aga	aga	agg	tat	caa	a										490
Arg	Arg	Arg	Arg	Tyr	Gln											
			75													

<210> 592

<211> 537

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 197..535

<220>

<221> sig_peptide

<222> 197..304

<223> Von Heijne matrix

score 3.79999995231628

seq WLGCIIVFPFSALL/QQ

<400> 592

ccttttttaaa	caatatccat	tctcttgatg	tctacaatat	caccttttctt	atagattcgc	60										
atatacatgg	ccaaaggaac	aactccatgt	tttctaaaag	gcctagagaa	catacatcag	120										
ctcaccagga	catgattata	gcttctcaag	cttatctaaa	ataaaagata	acatatatat	180										
caacattttt	gatgaa	atg	atg	act	gaa	aaa	cat	gag	gat	cac	tgt	ctc	aag	232		
		Met	Met	Thr	Glu	Lys	His	Glu	Asp	His	Cys	Leu	Lys			
		-35					-30					-25				
agc	tgt	agt	ggt	cac	tca	tat	ata	aga	aag	aat	tgg	ctt	gga	tgc	att	280
Ser	Cys	Ser	Gly	His	Ser	Tyr	Ile	Arg	Lys	Asn	Trp	Leu	Gly	Cys	Ile	
			-20				-15					-10				
gtc	ttc	cct	ttt	tct	gct	ctt	ctg	caa	caa	tct	gag	att	agt	gga	aca	328
Val	Phe	Pro	Phe	Ser	Ala	Leu	Leu	Gln	Gln	Ser	Glu	Ile	Ser	Gly	Thr	
			-5				1				5					
ttt	caa	gta	act	att	cca	cca	gtt	ttg	ctc	ggg	tat	act	tgg	agt	aat	376
Phe	Gln	Val	Thr	Ile	Pro	Pro	Val	Leu	Leu	Gly	Tyr	Thr	Trp	Ser	Asn	
	10					15				20						
act	tat	gta	ttt	cct	aaa	gaa	gat	tcc	aat	gag	cag	aat	tta	aaa	gaa	424
Thr	Tyr	Val	Phe	Pro	Lys	Glu	Asp	Ser	Asn	Glu	Gln	Asn	Leu	Lys	Glu	
25				30				35						40		
tgt	aca	ttc	tta	aat	att	ttt	gct	acc	att	gaa	cct	caa	ata	tca	tat	472
Cys	Thr	Phe	Leu	Asn	Ile	Phe	Ala	Thr	Ile	Glu	Pro	Gln	Ile	Ser	Tyr	
			45				50						55			
gtc	acc	tgt	aat	cca	aca	cta	gat	aag	ttt	tta	gat	caa	aca	gag	gtc	520
Val	Thr	Cys	Asn	Pro	Thr	Leu	Asp	Lys	Phe	Leu	Asp	Gln	Thr	Glu	Val	
			60				65					70				
ttg	cag	aga	gca	cag	at											537
Leu	Gln	Arg	Ala	Gln												
			75													

<210> 593

<211> 488

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 79..486

<220>

<221> sig_peptide

<222> 79..201

<223> Von Heijne matrix

score 3.59999990463257

seq LHTSVTLFLLSYC/DC

<400> 593

```

agatgattcc ctgattctcc agagagatta cacacttcgt ttgtggctaa ggttactgtg      60
accaatgaa agaagaaa atg aaa gcc ata aag aaa agt ctt aca gaa gaa      111
                        Met Lys Ala Ile Lys Lys Ser Leu Thr Glu Glu
                        -40                                -35
gaa tac ctg tac ctg gac ttt tct cac caa aca gaa gga tgc atc ttt      159
Glu Tyr Leu Tyr Leu Asp Phe Ser His Gln Thr Glu Gly Cys Ile Phe
-30                                -25                                -20                                -15
cct ctt cat aca tct gta act tta ttt ctg tta tct tac tgt gac tgt      207
Pro Leu His Thr Ser Val Thr Leu Phe Leu Leu Ser Tyr Cys Asp Cys
                        -10                                -5                                1
aaa atc ttt aaa att tgc tta gtt gtc acc aaa gag gtg agt aga gat      255
Lys Ile Phe Lys Ile Cys Leu Val Val Thr Lys Glu Val Ser Arg Asp
                        5                                10                                15
agt tca cta cta aga gat gac ctg atc cag gat gtt gaa ata cag att      303
Ser Ser Leu Leu Arg Asp Asp Leu Ile Gln Asp Val Glu Ile Gln Ile
20                                25                                30
att tca agg cag gag ctc cca cca ata gtc caa aat tgc tgt ttg cct      351
Ile Ser Arg Gln Glu Leu Pro Pro Ile Val Gln Asn Cys Cys Leu Pro
35                                40                                45                                50
gca gta gta gaa cga tca gac aat ttt tgt aga gca gga ctt gct gtt      399
Ala Val Val Glu Arg Ser Asp Asn Phe Cys Arg Ala Gly Leu Ala Val
55                                60                                65
gta ttg aga cac ata atc cag aaa tcc tat gaa gca gac ccc tta aag      447
Val Leu Arg His Ile Ile Gln Lys Ser Tyr Glu Ala Asp Pro Leu Lys
70                                75                                80
aag gaa ctt ttg gaa ctt ctg ggc ttw aaa aga ctt gct tg      488
Lys Glu Leu Leu Glu Leu Leu Gly Xaa Lys Arg Leu Ala
85                                90                                95

```

<210> 594

<211> 490

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 80..490

<220>

<221> sig_peptide

<222> 80..127

<223> Von Heijne matrix

score 3.59999990463257

seq FSLLSISGPPISS/SA

<400> 594

```

gaatgtttat cctctggaca aaccagccag cctctccaga gcaggcgtgt gatctctgta      60
ccccgcgagt ggtcagaat atg gag aac ttc tca ctc ctc agc atc tct gga      112
                        Met Glu Asn Phe Ser Leu Leu Ser Ile Ser Gly

```

324

																-15																	-10																
cct	cca	atc	tct	tcc	tcc	gcc	ctg	agt	gct	ttt	ccc	gac	att	atg	ttc		160																																
Pro	Pro	Ile	Ser	Ser	Ser	Ala	Leu	Ser	Ala	Phe	Pro	Asp	Ile	Met	Phe																																		
-5																5	10																																
tct	cg	gcc	acc	agc	ctg	cca	gac	att	gca	aag	aca	gca	gta	ccc	act		208																																
Ser	Arg	Ala	Thr	Ser	Leu	Pro	Asp	Ile	Ala	Lys	Thr	Ala	Val	Pro	Thr																																		
15																20	25																																
gag	gca	tcc	agc	cca	gct	cag	gcc	ctg	cca	ccc	cag	tac	caa	agc	atc		256																																
Glu	Ala	Ser	Ser	Pro	Ala	Gln	Ala	Leu	Pro	Pro	Gln	Tyr	Gln	Ser	Ile																																		
30																35	40																																
att	gtc	agg	caa	ggg	ata	cag	aac	aca	gtg	ctc	tca	cca	grc	wgc	agc		304																																
Ile	Val	Arg	Gln	Gly	Ile	Gln	Asn	Thr	Val	Leu	Ser	Pro	Xaa	Xaa	Ser																																		
45																50	55																																
ttg	ggg	gac	acc	cag	cac	gga	gag	aag	ctg	agg	cgg	aac	tgc	act	atc		352																																
Leu	Gly	Asp	Thr	Gln	His	Gly	Glu	Lys	Leu	Arg	Arg	Asn	Cys	Thr	Ile																																		
60																65	70																																
tac	cg	ccc	tgg	ttc	tcc	ccc	tac	agc	tac	ttc	gtg	tgt	gca	gac	aaa		400																																
Tyr	Arg	Pro	Trp	Phe	Ser	Pro	Tyr	Ser	Tyr	Phe	Val	Cys	Ala	Asp	Lys																																		
80																85	90																																
gag	agc	cag	ctg	gag	gcc	tat	gac	ttc	cca	gag	gtg	cag	cag	gat	gag		448																																
Glu	Ser	Gln	Leu	Glu	Ala	Tyr	Asp	Phe	Pro	Glu	Val	Gln	Gln	Asp	Glu																																		
95																100	105																																
ggc	aag	tgg	gac	aac	tgc	ctt	tct	gag	gac	atg	gct	gag	aac				490																																
Gly	Lys	Trp	Asp	Asn	Cys	Leu	Ser	Glu	Asp	Met	Ala	Glu	Asn																																				
110																115	120																																

```
<210> 595
<211> 495
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 157..495
```

<400> 595																
attcttttcga		caggagttga		aaccagcacc		ttcccttttct		ctgagtcctg		cctccttctg						60
cagaaggggag		ctcaaaagaa		ctttgtttgtt		ttgcctttta		ctctgggggtg		aaagcggcag						120
gaggtatgtg		agatggtgaa		atgatttgct		tctgcc		atg		ctg ggg tca		cgg gtg				174
								Met		Leu Gly Ser		Arg Val				
								1				5				
gat cgc cct		aaa ctc		tcg gtg		gcc ccc		tca gta		gtt ttg		gaa gag		gac		222
Asp Arg Pro		Lys Leu		Ser Val		Ala Pro		Ser Val		Val Val		Leu Glu		Glu Asp		
		10				15						20				
caa gtc ctt		gtc tct		cca gca		gtg gac		ctg gaa		gma gga		tgc cgg		ctc		270
Gln Val Leu		Val Ser		Pro Ala		Val Asp		Leu Glu		Xaa Gly		Cys Arg		Leu		
		25				30						35				
agg gac ttc		act gag		aaa ata		atg aat		gtc aaa		gga aaa		gta att		ctg		318
Arg Asp Phe		Thr Glu		Lys Ile		Met Asn		Val Lys		Gly Gly		Lys Val		Ile Leu		
		40				45				50						
tca atg ctg		gtt gtc		tca act		gtg atc		att gtg		ttt tgg		gaa ttt		atc		366
Ser Met Leu		Val Val		Ser Thr		Val Ile		Ile Val		Phe Trp		Glu Phe		Ile		
55				60				65				70				
aac agc aca		gaa ggc		tct ttc		ttg tgg		ata tat		cac tca		aaa aac		cca		414
Asn Ser Thr		Glu Gly		Ser Phe		Leu Trp		Ile Tyr		His Ser		Lys Asn		Pro		
		75				80						85				
gaa gtt gat		gac agc		agt gct		cag aag		ggc tgg		tgg ttt		ctg agc		tgg		462
Glu Val Asp		Asp Ser		Ser Ala		Gln Lys		Gly Trp		Trp Phe		Leu Ser		Trp		
		90				95						100				
ttt aac aat		ggg atc		cac aat		tat caa		caa ggg								495
Phe Asn Asn		Gly Ile		His Asn		Tyr Gln		Gln Gly								
		105				110										

<210> 596
 <211> 939
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 416..937

<400> 596
 aaaaggagta gctattagcc aattcggcag ggcccgcttt ttagaagctt gatttctyttt 60
 gaagatgaaa gactagcgga agctctgcct ctttccccag tggcgaggga actcggggcg 120
 attggctggg aactgtatcc acccaaagt caccgatttc ttcctatgca ggaaatgagc 180
 agacccatca ataagaaatt tctcagcctg gccgaaaatg gttggcccca cgaagccacg 240
 acaactggag gcaaagaggg ttgctcaacg ccccgccctca ttggaaaacc aaatcagatc 300
 tgggacctat atagcgtggc ggaggcgggg cgatgattgt cgcgctcgca cccactgcag 360
 ctgcgccacag tcgcatttct ttccccgccc ctgagaccct gcagcaccat ctgtc atg 418
 Met
 1
 gcg gct ggg ctg ttt ggt ttg agc gct cgc cgt ctt ttg gcg gca gcg 466
 Ala Ala Gly Leu Phe Gly Leu Ser Ala Arg Arg Leu Leu Ala Ala Ala
 5 10 15
 gcg acg cga ggg ctc ccg gcc gcc cgc gtc cgc tgg gaa tct agc ttc 514
 Ala Thr Arg Gly Leu Pro Ala Ala Arg Val Arg Trp Glu Ser Ser Phe
 20 25 30
 tcc agg act gtg gtc gcc ccg tcc gct gtg gcg gga aag cgg ccc cca 562
 Ser Arg Thr Val Val Ala Pro Ser Ala Val Ala Gly Lys Arg Pro Pro
 35 40 45
 gaa ccg acc aca ccg tgg caa gag gac cca gaa ccc gag gac gaa aac 610
 Glu Pro Thr Thr Pro Trp Gln Glu Asp Pro Glu Pro Glu Asp Glu Asn
 50 55 60 65
 ttg tat gag aag aac cca gac tcc cat ggt tat gac aag gac ccc gtt 658
 Leu Tyr Glu Lys Asn Pro Asp Ser His Gly Tyr Asp Lys Asp Pro Val
 70 75 80
 ttg gac gtc tgg aac atg cga ctt gtc ttc ttc ttt ggc gtc tcc atc 706
 Leu Asp Val Trp Asn Met Arg Leu Val Phe Phe Phe Gly Val Ser Ile
 85 90 95
 atc ctg gtc ctt ggc agc acc ttt gtg gcc tat ctg cct gac tac agg 754
 Ile Leu Val Leu Gly Ser Thr Phe Val Ala Tyr Leu Pro Asp Tyr Arg
 100 105 110
 atg aaa gag tgg tcc cgc cgc gaa gct gag agg ctt gtg aaa tac cga 802
 Met Lys Glu Trp Ser Arg Arg Glu Ala Glu Arg Leu Val Lys Tyr Arg
 115 120 125
 gag gcc aat ggc ttc cca tca tgg aat cca act gct tcg acc cca gca 850
 Glu Ala Asn Gly Phe Pro Ser Trp Asn Pro Thr Ala Ser Thr Pro Ala
 130 135 140 145
 aga tcc agc tgc cag agg atg agt gac cag ttg cta agt ggg gct caa 898
 Arg Ser Ser Cys Gln Arg Met Ser Asp Gln Leu Leu Ser Gly Ala Gln
 150 155 160
 gaa gca ccg cct tcc cca ccc cct gcc tgc cat tct gac ct 939
 Glu Ala Pro Pro Ser Pro Pro Pro Ala Cys His Ser Asp
 165 170

<210> 597
 <211> 528
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 51..527

<400> 597

326

```

aagctggggg agctcgtcgc cgccgccggc ggctagcggg cgtccgcgcc atg gag      56
                                     Met Glu
                                     1
cgc tac gcg gcc gcc ttg gag gag gtg gcg gac ggt gcc cgg cag cag      104
Arg Tyr Ala Ala Ala Leu Glu Glu Val Ala Asp Gly Ala Arg Gln Gln
      5      10      15
gag cga cac tac cag ttg ctg tcg gcg cta cag agc ctg gtg aag gag      152
Glu Arg His Tyr Gln Leu Leu Ser Ala Leu Gln Ser Leu Val Lys Glu
      20      25      30
ttg ccc agc tct ttc cag cag cgc ctg tcc tac acc acg ctc agc gac      200
Leu Pro Ser Ser Phe Gln Gln Arg Leu Ser Tyr Thr Thr Leu Ser Asp
      35      40      45      50
ctg gcc ctg gcg ctt ctc gac ggc act gtg ttc gaa atc gtg cag ggg      248
Leu Ala Leu Ala Leu Leu Asp Gly Thr Val Phe Glu Ile Val Gln Gly
      55      60      65
cta ctg gag atc cag cac ctc acc gaa aag agc ctg tac aac cag cgc      296
Leu Leu Glu Ile Gln His Leu Thr Glu Lys Ser Leu Tyr Asn Gln Arg
      70      75      80
ctg cgc cta cag aac gag cac cga gtg ctc agg cag gcg ctg cgg cag      344
Leu Arg Leu Gln Asn Glu His Arg Val Leu Arg Gln Ala Leu Arg Gln
      85      90      95
aag cac cag gaa gcc cag cag gcc tgc cgg ccc cac aac ctg cct gtg      392
Lys His Gln Glu Ala Gln Gln Ala Cys Arg Pro His Asn Leu Pro Val
      100      105      110
gtt cag gcg gct cag cag cga gaa cta gag gcc gtg gaa cac cgg atc      440
Val Gln Ala Ala Gln Gln Arg Glu Leu Glu Ala Val Glu His Arg Ile
      115      120      125      130
cgt gag gag cag cgg gcg atg gac cat aag atc atc ctg gag ctg gac      488
Arg Glu Glu Gln Arg Ala Met Asp His Lys Ile Ile Leu Glu Leu Asp
      135      140      145
cgg aag gtg gct gac cag cag agc aca ctg gag aag gcg g      528
Arg Lys Val Ala Asp Gln Gln Ser Thr Leu Glu Lys Ala
      150      155

```

<210> 598

<211> 556

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 128..556

<400> 598

```

cttctttggc cctgtgacac gtagcaacgg ggctggttca gggctctgaaa cagagtttgg      60
gggttggttg ggattagtga agctactgcc ttgcccga gcgcasctca gagtttgatt      120
atttgca atg tca ggc ttt gaa aac tta aac acg gat ttc tac cag aca      169
      Met Ser Gly Phe Glu Asn Leu Asn Thr Asp Phe Tyr Gln Thr
      1      5      10
agt tac agc atc gat gat cag tca cag cag tcc tat gat tat gga gga      217
Ser Tyr Ser Ile Asp Asp Gln Ser Gln Gln Ser Tyr Asp Tyr Gly Gly
      15      20      25      30
agt gga gga ccc tat agc aaa cag tat gct ggc tat gac tat tcg cag      265
Ser Gly Gly Pro Tyr Ser Lys Gln Tyr Ala Gly Tyr Asp Tyr Ser Gln
      35      40      45
caa ggc aga ttt gtc cct cca gac atg atg cag cca caa cag cca tac      313
Gln Gly Arg Phe Val Pro Pro Asp Met Met Gln Pro Gln Gln Pro Tyr
      50      55      60
acc ggg cag att tac cag cca act cag gca tat act cca gct tca cct      361
Thr Gly Gln Ile Tyr Gln Pro Thr Gln Ala Tyr Thr Pro Ala Ser Pro
      65      70      75
cag cct ttc tat gga aac aac ttt gag gat gag cca cct tta tta gaa      409
Gln Pro Phe Tyr Gly Asn Asn Phe Glu Asp Glu Pro Pro Leu Leu Glu

```

327

80	85	90	
gag tta ggt atc aat ttt gac cac atc tgg caa aaa aca cta aca gta			457
Glu Leu Gly Ile Asn Phe Asp His Ile Trp Gln Lys Thr Leu Thr Val			
95	100	105	110
tta cat ccg tta aaa gta gca gat ggc agc atc atg aat gaa act gat			505
Leu His Pro Leu Lys Val Ala Asp Gly Ser Ile Met Asn Glu Thr Asp			
	115	120	125
ttg gca ggt cca atg gtt ttt tgc ctt gct ttt tgg agc aca ttg cta			553
Leu Ala Gly Pro Met Val Phe Cys Leu Ala Phe Trp Ser Thr Leu Leu			
	130	135	140
ctg			556
Leu			

<210> 599
 <211> 564
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 151..564

<400> 599
 ataacaaaac aatccaacta tactgaataa tacagaatca aactgggatc caaaaacaat 60
 ttaaaacaaa gttggaaact caaccgaaac aagaacaaat cgaatccgac atgaaacaag 120
 tctcacaagt gttcttaagc tgacagaaaag atg gtc aaa acc aaa ctt aaa acc 174
 Met Val Lys Thr Lys Leu Lys Thr

gta gta atc gac ata acg gtc ata cca ttc gcc gac gat ttt ttc tgc	222			
Val Val Ile Asp Ile Thr Val Ile Pro Phe Ala Asp Asp Phe Phe Cys				
10	15	20		
cct gga gct cca gcc ggc gta gag ata tca cca gcg gct gat gga cca	270			
Pro Gly Ala Pro Ala Gly Val Glu Ile Ser Pro Ala Ala Asp Gly Pro				
25	30	35	40	
tct cct gaa gga ccc atc gcc gag cct gac gga gaa tct gac ggt gaa	318			
Ser Pro Glu Gly Pro Ile Ala Glu Pro Asp Gly Glu Ser Asp Gly Glu				
	45	50	55	
gaa gaa gta gag cct ccc gga gcc aac gat gga gaa acg gtg ggt gag	366			
Glu Glu Val Glu Pro Pro Gly Ala Asn Asp Gly Glu Thr Val Gly Glu				
	60	65	70	
ttt gaa aca ggg gat gaa gat tta ggt gcc att gat cca ggt ggt gaa	414			
Phe Glu Thr Gly Asp Glu Asp Leu Gly Ala Ile Asp Pro Gly Gly Glu				
	75	80	85	
gtc atc gga gca gaa gac gga gga ata gtt gaa gat gat tta ggt gcc	462			
Val Ile Gly Ala Glu Asp Gly Gly Ile Val Glu Asp Asp Leu Gly Ala				
	90	95	100	
att ggt gcc ggc ggt gaa gtc atc gga gca gag gat ggg gaa aca ggg	510			
Ile Gly Ala Gly Gly Glu Val Ile Gly Ala Glu Asp Gly Glu Thr Gly				
	105	110	115	120
gac gtg gat tta ggt ggt gcc ggt ggt gaa gtc gtc gga gaa aca ggg	558			
Asp Val Asp Leu Gly Gly Ala Gly Gly Glu Val Val Gly Glu Thr Gly				
	125	130	135	
gag ctc			564	
Glu Leu				

<210> 600
 <211> 421
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 198..419

<220>

<221> misc_feature

<222> 275..278

<223> n=a, g, c or t

<400> 600

```

cttttacttc aggcgaacct gaactcagtt aatgcaattt tgaagcatgc cctgaataga      60
ttcagtcatt atcaagtcaa actaaaaacg gtgaaagggt gcgactatta ccaaataagga      120
aaaacctgaa gacataagaa ctacacatga ggaatatgtc atttagcact ttcacttttt      180
gatctccaca gaagaca atg aga agt cat acc ata aca atg acg aca act      230
                Met Arg Ser His Thr Ile Thr Met Thr Thr Thr
                1             5             10
tca gtc agc agc tgg cct tac tcc tcc cac aga atg cgc ttt awn nnn      278
Ser Val Ser Ser Trp Pro Tyr Ser Ser His Arg Met Arg Phe Xaa Xaa
                15             20             25
hat cat agc gac caa ccg cca caa aac ttc tca gca aca cca aat gtt      326
Xaa His Ser Asp Gln Pro Pro Gln Asn Phe Ser Ala Thr Pro Asn Val
                30             35             40
act acc tgt ccc atg gat gaa aaa ttg khs act act gtg tta acc aca      374
Thr Thr Cys Pro Met Asp Glu Lys Leu Xaa Thr Thr Val Leu Thr Thr
                45             50             55
tcc tac tct gtt att ttc atc gtg gga ctg gtt ggg aac ata atc gc      421
Ser Tyr Ser Val Ile Phe Ile Val Gly Leu Val Gly Asn Ile Ile
60                65                70

```

<210> 601

<211> 529

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 124..528

<400> 601

```

ggctgcgctt agtgggcccgt tgccttacag ttgctgagag gaggcgagag gcggggggcgc      60
tagggccgag atcatgtctg actgggagag gtttccttgg cagcagagga cgctagggtt      120
ggg atg aaa gaa gct ggg cag atg saa aat ctg gag agc gcg agg gcc      168
    Met Lys Glu Ala Gly Gln Met Xaa Asn Leu Glu Ser Ala Arg Ala
    1             5             10             15
ggg cgg tca gtc agc acc cag act ggc agc atg acc ggt cag ata cca      216
Gly Arg Ser Val Ser Thr Gln Thr Gly Ser Met Thr Gly Gln Ile Pro
                20             25             30
agg ctt tct aaa gtc aac ctt ttc act ctg ctc agc ctc tgg atg gag      264
Arg Leu Ser Lys Val Asn Leu Phe Thr Leu Leu Ser Leu Trp Met Glu
                35             40             45
ctc ttt cca gca gaa gcc cag cgg caa aaa tct cag aaa aat gaa grg      312
Leu Phe Pro Ala Glu Ala Gln Arg Gln Lys Ser Gln Lys Asn Glu Xaa
                50             55             60
gga aag cat gga ccc tta gga gat aat gaa gag agg acc aga gta tct      360
Gly Lys His Gly Pro Leu Gly Asp Asn Glu Glu Arg Thr Arg Val Ser
                65             70             75
act gac aaa aga cag gta aag aga act ggt ctt gtg gtg gta aaa aac      408
Thr Asp Lys Arg Gln Val Lys Arg Thr Gly Leu Val Val Val Lys Asn
80                85                90                95
atg aaa att gtt ggt ctc cac tgt tct agt gaa gat tta cat gcc ggg      456
Met Lys Ile Val Gly Leu His Cys Ser Ser Glu Asp Leu His Ala Gly
                100             105             110
cag att gct ctt att aaa cat ggg tca agg ctg aaa aac tgt gat cct      504
Gln Ile Ala Leu Ile Lys His Gly Ser Arg Leu Lys Asn Cys Asp Pro
                115             120             125
tat ttt tcc aga aaa cmt gtt ctg c      529

```

Tyr Phe Ser Arg Lys Xaa Val Leu
130 135

<210> 602
<211> 208
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 24..206

<400> 602
ccttccctgc ggccgcgcag agc atg ccg gga gcc gcc ggc cgc cgg tgt ccg 53
Met Pro Gly Ala Ala Gly Arg Arg Cys Pro
1 5 10
gac cgc tcg ccc ccg ttt gga ccc gac ttc ggt tct tct ggg gtg ttg 101
Asp Arg Ser Pro Pro Phe Gly Pro Asp Phe Gly Ser Ser Gly Val Leu
15 20 25
atg ctc cta aag ccc gag agc acg tgt cca gac cct agc ctg tac gac 149
Met Leu Leu Lys Pro Glu Ser Thr Cys Pro Asp Pro Ser Leu Tyr Asp
30 35 40
gcd baa ctc tgc ccg gtc cca gaa cca aag cca tgc cgg ggt gtg gcc 197
Ala Xaa Leu Cys Pro Val Pro Glu Pro Lys Pro Cys Arg Gly Val Ala
45 50 55
tct gac ccc ac 208
Ser Asp Pro
60

<210> 603
<211> 587
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 198..587

<400> 603
cttttacttc aggcgaacct gaactcagtt aatgcaatth tgaagcatgc cctgaataga 60
ttcagtcatt atcaagtcaa actaaaaacg gtgaaagggt gcgactatta ccaaataagga 120
aaaacctgaa gacataagaa ctacacatga ggaatatgtc atttagcact ttcacttttt 180
gatctccaca gaagaca atg aga agt cat acc ata aca atg acg aca act 230
Met Arg Ser His Thr Ile Thr Met Thr Thr
1 5 10
tca gtc agc agc tgg cct tac tcc tcc cac aga atg cgc ttt ata acc 278
Ser Val Ser Ser Trp Pro Tyr Ser Ser His Arg Met Arg Phe Ile Thr
15 20 25
aat cat agc gac caa ccg cca caa aac ttc tca gca aca cca aat gtt 326
Asn His Ser Asp Gln Pro Pro Gln Asn Phe Ser Ala Thr Pro Asn Val
30 35 40
act acc tgt ccc atg gat gaa aaa ttg cta tct act gtg tta acc aca 374
Thr Thr Cys Pro Met Asp Glu Lys Leu Leu Ser Thr Val Leu Thr Thr
45 50 55
tcc tac tct gtt att ttc atc gtg gga ctg gtt ggg aac ata atc gcc 422
Ser Tyr Ser Val Ile Phe Ile Val Gly Leu Val Gly Asn Ile Ile Ala
60 65 70 75
ctc tat gta ttt ctg ggt att cac cgt aaa aga aat tcc att caa att 470
Leu Tyr Val Phe Leu Gly Ile His Arg Lys Arg Asn Ser Ile Gln Ile
80 85 90
tat cta ctt aac gta gcc att gca gac ctc cta ctc atc ttc tgc ctc 518
Tyr Leu Leu Asn Val Ala Ile Ala Asp Leu Leu Leu Ile Phe Cys Leu
95 100 105

330

```

cct ttc cga ata atg tat cat att aac caa aac aag tgg aca cta ggt      566
Pro Phe Arg Ile Met Tyr His Ile Asn Gln Asn Lys Trp Thr Leu Gly
      110                      115                      120
gtg att ytg tgc aag gtt gtg      587
Val Ile Leu Cys Lys Val Val
      125                      130

```

<210> 604
 <211> 528
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 327..527

```

<400> 604
acaagaaaag aacatggtct agactgaagt accaactaaa tcattctcctt tcaaattatc      60
accgacacca tcattggattc aagcaccgca cacagtccgg tgtttctggt atttcctcca      120
gaaatcactg cttcagaata tgagtccaca gaactttcag ccacgacctt ttcaactcaa      180
agccccctgc aaaaattatt tgctagaaaa atgaaaatct taggggatat ccattctggg      240
gctctgtttt gttcattaat tctggagcct tcctaattgc agtgaaaaga aaaaccacag      300
aaactctgat aatattgagc cgaata atg aat ttt ctt agt gcc ctg gga gca      353
                        Met Asn Phe Leu Ser Ala Leu Gly Ala
                        1                      5
ata gct gga atc att ctc ctc aca ttt ggt ttc atc cta gat caa aac      401
Ile Ala Gly Ile Ile Leu Leu Thr Phe Gly Phe Ile Leu Asp Gln Asn
10                      15                      20                      25
tac att tgt ggt tat tct cac caa aat agt cag tgt aag gct gtt act      449
Tyr Ile Cys Gly Tyr Ser His Gln Asn Ser Gln Cys Lys Ala Val Thr
      30                      35                      40
gtc ctg ttc ttg gga att ttg att aca ttg atg act ttc agc att att      497
Val Leu Phe Leu Gly Ile Leu Ile Thr Leu Met Thr Phe Ser Ile Ile
      45                      50                      55
gaa tta ttc att tct ctg ctt tct caa ttt t      528
Glu Leu Phe Ile Ser Leu Leu Ser Gln Phe
      60                      65

```

<210> 605
 <211> 227
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 62..226

```

<400> 605
atttcgggtt ccggtgtcag ttcgaggcgc cgccgcccgc gccgcagccg ccggagccgc      60
a atg cct aaa gga gga aga aag gga ggc cac aaa ggc cgg gcg agg cag      109
Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg Gln
1                      5                      10                      15
tat aca agc cct gag gag atc gac gcg cas tgc agg ctg aga agc aga      157
Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser Arg
      20                      25                      30
agg cca ggg aag aag agg agc aaa aaa ggt gga gat ggg gct gca ggt      205
Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala Gly
      35                      40                      45
gac ccc aaa aaa aaa aaa aaa a      227
Asp Pro Lys Lys Lys Lys Lys
      50                      55

```

<210> 606

<211> 228
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 63..227

<400> 606
 agaacatcct aatcrwaaaa gttcgaggcg ccgcccgcgc cgccgcagcc gccggagccg 60
 ca atg cct aaa gga gga aga aag gga ggc cac aaa ggc cgg gcg agg 107
 Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg
 1 5 10 15
 cag tat aca agc cct gag gag atc gac gcg cas tgc agg ctg aga agc 155
 Gln Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser
 20 25 30
 aga agg cca ggg aag aag agg agc aaa aaa ggt gga gat ggg gct gca 203
 Arg Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala
 35 40 45
 ggt gac ccc aaa aaa aaa aaa a 228
 Gly Asp Pro Lys Lys Lys Lys Lys
 50 55

<210> 607
 <211> 378
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 119..376

<220>
 <221> misc_feature
 <222> 143,318,332
 <223> n=a, g, c or t

<400> 607
 gaaagacgga gagaggaaaa gggacaaggg caaggcaaga gagaggcgctc gtcgtttgat 60
 ctgctgccgg agtgcggaaga ggaagaggaa gaggaagagg agcgccggag ttgtaaag 118
 atg tts ctg gtg gac tgg ttc tat ngg gtg ctg tca tcg ctt ggg ctg 166
 Met Xaa Leu Val Asp Trp Phe Tyr Xaa Val Leu Ser Ser Leu Gly Leu
 1 5 10 15
 tgg cag aag gag gct aag atc ctc ttc ctt ggc ctc gac aac gcc ggc 214
 Trp Gln Lys Glu Ala Lys Ile Leu Phe Leu Gly Leu Asp Asn Ala Gly
 20 25 30
 aag acc acc ctc ctc cac atg ctg aag gac gag cgg ctc gta cag cac 262
 Lys Thr Thr Leu Leu His Met Leu Lys Asp Glu Arg Leu Val Gln His
 35 40 45
 cag cca acg cag tac ccc acg tca gaa gag ttg agc atc ggc agg atc 310
 Gln Pro Thr Gln Tyr Pro Thr Ser Glu Glu Leu Ser Ile Gly Arg Ile
 50 55 60
 aag ttc ang gcg ttc gac ctt ngg ggc cac cag atc gcc cgc cgc gtc 358
 Lys Phe Xaa Ala Phe Asp Leu Xaa Gly His Gln Ile Ala Arg Arg Val
 65 70 75 80
 tgg aag gac tac tmc gcc aa 378
 Trp Lys Asp Tyr Xaa Ala
 85

<210> 608
 <211> 455
 <212> DNA
 <213> Homo sapiens

<220>

<221> CDS

<222> 63..455

<400> 608

```

attscacact gtgatgctca gagagaaacc atgagccctg aaaaacacca tattgttttg      60
tg atg tca cct ctg gaa tgt tct gag tgt ttt ggt gac caa ctt ctg      107
  Met Ser Pro Leu Glu Cys Ser Glu Cys Phe Gly Asp Gln Leu Leu
    1          5          10          15
cat agg acc tat acc tgg caa ctc aca ttg cac tca agg cca aat tat      155
His Arg Thr Tyr Thr Trp Gln Leu Thr Leu His Ser Arg Pro Asn Tyr
          20          25          30
aca aga aaa aga gat acc aga tct gaa agc cta gaa att cca atc agt      203
Thr Arg Lys Arg Asp Thr Arg Ser Glu Ser Leu Glu Ile Pro Ile Ser
          35          40          45
gtg gtt cta cct cag agg ggc aca gcc gaa ccc ttc ccg agg ctc cac      251
Val Val Leu Pro Gln Arg Gly Thr Ala Glu Pro Phe Pro Arg Leu His
          50          55          60
aac ttg tac agc acc cct cgc tgc gcg cag cag gcc gcc ctg ccc cgg      299
Asn Leu Tyr Ser Thr Pro Arg Cys Ala Gln Gln Ala Ala Leu Pro Arg
          65          70          75
ctg agc cgc agg atg gcg agc cag cac tcc tat cca ctg aac cgc ttc      347
Leu Ser Arg Arg Met Ala Ser Gln His Ser Tyr Pro Leu Asn Arg Phe
          80          85          90          95
tcc tcc gtg cct tta gac ccc atg gag cgc ccc atg tcc cag gcc gac      395
Ser Ser Val Pro Leu Asp Pro Met Glu Arg Pro Met Ser Gln Ala Asp
          100          105          110
ctg gag ctg gac tac aac ccg ccg cgg gtg cag ctc agc gac gag atg      443
Leu Glu Leu Asp Tyr Asn Pro Pro Arg Val Gln Leu Ser Asp Glu Met
          115          120          125
ttc ktg ttc car
Phe Xaa Phe Gln
          130

```

<210> 609

<211> 366

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 68..364

<400> 609

```

gtacagaaag attggtggac aggagcagcg gccggtgggm gagggcgctc ggcggcggcc      60
tgcggcc atg gcc acc gtg atg gca gcg acg gcg gcg gag cgg gcg gtg      109
  Met Ala Thr Val Met Ala Ala Thr Ala Ala Glu Arg Ala Val
    1          5          10
ctg gag gag gag ttc cgc tgg ctg ctg cac gac gak ktg cac gct gtg      157
Leu Glu Glu Glu Phe Arg Trp Leu Leu His Asp Xaa Xaa His Ala Val
    15          20          25          30
ttg wwg cag ctg cag gac atc ctc aag gag gcc tct ctg cgc ttc act      205
Leu Xaa Gln Leu Gln Asp Ile Leu Lys Glu Ala Ser Leu Arg Phe Thr
          35          40          45
ctg ccg gcc tcc ggc act gag ggg ccc gcc aag caa gag aac ttc atc      253
Leu Pro Gly Ser Gly Thr Glu Gly Pro Ala Lys Gln Glu Asn Phe Ile
          50          55          60
cta ggc agc tgt ggc aca gac cak gtg aag ggt gtg ctg act ctg cak      301
Leu Gly Ser Cys Gly Thr Asp Xaa Val Lys Gly Val Leu Thr Leu Xaa
          65          70          75
ggg gat gcc ctc agc cag gcg gat gtg aac ctg aag atg ccc cgg aac      349
Gly Asp Ala Leu Ser Gln Ala Asp Val Asn Leu Lys Met Pro Arg Asn

```

80 85 90 366
 aac cag ctg ctg cac tt
 Asn Gln Leu Leu His
 95

<210> 610
 <211> 498
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 209..496

<400> 610
 atgatgccag tgaatgcgta gtcccagcct cagcacagg gagccacctt gaagctctca 60
 aatatcactg ttgtgaatac agagagggaa aaccaactgt aacgtgccac ccaaataaag 120
 ctcatcacta gtccctgtcca gcaacgtgcc tctcctggcc ctagagttct tggaatagc 180
 ccaggccaaa gagaaggcct ttctcccc atg gtc agc cac acg ttc cac atg 232
 Met Val Ser His Thr Phe His Met
 1 5
 cgc aca gag gag tct gat gcc tca cag gag ggc gat gac cta ccc aag 280
 Arg Thr Glu Glu Ser Asp Ala Ser Gln Glu Gly Asp Asp Leu Pro Lys
 10 15 20
 tcc tca gca aac acc agc cat ccc aag cag gat gac agc ccc aag tcc 328
 Ser Ser Ala Asn Thr Ser His Pro Lys Gln Asp Asp Ser Pro Lys Ser
 25 30 35 40
 tca gaa gaa acc atc cag ccc aag gag ggt gac atc ccc aag gcc cca 376
 Ser Glu Glu Thr Ile Gln Pro Lys Glu Gly Asp Ile Pro Lys Ala Pro
 45 50 55
 gaa gaa acc atc caa tcc aag aag gag gac ctc ccc aag tcc tca gaa 424
 Glu Glu Thr Ile Gln Ser Lys Lys Glu Asp Leu Pro Lys Ser Ser Glu
 60 65 70
 aaa gcc atc cag ccc aaa gag agt aac atc ccc aag tcc tca gca aaa 472
 Lys Ala Ile Gln Pro Lys Glu Ser Asn Ile Pro Lys Ser Ser Ala Lys
 75 80 85
 ccc atc cag ccc aag ctg gca ata tt 498
 Pro Ile Gln Pro Lys Leu Ala Ile
 90 95

<210> 611
 <211> 504
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 214..504

<400> 611
 aagacagaca cggaagtgtc gggaggcgcc gggagcccgt tcggttgccg gtgtctctgg 60
 ccctgcggtc agccctggga acgtcccga gagctagatt cctagaggcc cgattccgct 120
 agcccgaac agacaaagcc agcgtcccgc cccgctcccc gacttaggat ccgatgccg 180
 cagcgtcctg gggcccccggt agcgggctgg acc atg agc ctg ctg gac gcc ctc 234
 Met Ser Leu Leu Asp Gly Leu
 1 5
 gct tcc tcg ccg cgg gct ccg ctg cag tcc agc aag gcc agg atg aaa 282
 Ala Ser Ser Pro Arg Ala Pro Leu Gln Ser Ser Lys Ala Arg Met Lys
 10 15 20
 aag ctc ccg aag aag agc cag aat gag aag tac cgg ctg aag tac ctg 330
 Lys Leu Pro Lys Lys Ser Gln Asn Glu Lys Tyr Arg Leu Lys Tyr Leu
 25 30 35
 cgg ctg cgc aaa gcg gcc aag gcc acg gtg ttt gaa aat gct gct att 378

334

Arg	Leu	Arg	Lys	Ala	Ala	Lys	Ala	Thr	Val	Phe	Glu	Asn	Ala	Ala	Ile	
40					45					50					55	
tgt	gat	gaa	att	gct	cgt	ctt	gag	gaa	aaa	ttt	ctt	aaa	gca	aaa	gaa	426
Cys	Asp	Glu	Ile	Ala	Arg	Leu	Glu	Glu	Lys	Phe	Leu	Lys	Ala	Lys	Glu	
				60					65					70		
gaa	aga	agg	tac	ttg	cta	aag	aag	ctc	ctc	cag	ctt	cag	gct	cta	act	474
Glu	Arg	Arg	Tyr	Leu	Leu	Lys	Lys	Leu	Leu	Gln	Leu	Gln	Ala	Leu	Thr	
			75					80					85			
gaa	ggg	gta	agt	aca	ggc	mtg	cag	ctc	ctt							504
Glu	Gly	Val	Ser	Thr	Gly	Xaa	Gln	Leu	Leu							
		90					95									

<210> 612
 <211> 483
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 161..481

<400> 612																
gtggg	cggtt	cggcg	gctct	gggcg	ctgtt	gtttg	gtctt	taggc	ctgcg	gaggg	gcgtt					60
atctg	gaggg	ccgcg	ggtgc	aggcc	gcagt	gacag	ggccg	ctcgc	ccccg	tagtc	ctgcc					120
tgtct	cccgg	tgcag	ctgtg	ttcgc	ggcct	gcagg	ccaac	atg	gcg	cag	gag	gtg				175
								Met	Ala	Gln	Glu	Val				
								1				5				
tcg	gag	tac	ctg	agc	cag	aac	ccg	cgg	gtg	gca	gcc	tgg	gtg	gag	gcg	223
Ser	Glu	Tyr	Leu	Ser	Gln	Asn	Pro	Arg	Val	Ala	Ala	Trp	Val	Glu	Ala	
				10				15						20		
ctg	cg	tgc	gac	ggc	gag	act	gac	aaa	cac	tgg	cg	cac	cg	cg	gat	271
Leu	Arg	Cys	Asp	Gly	Glu	Thr	Asp	Lys	His	Trp	Arg	His	Arg	Arg	Asp	
			25					30					35			
ttt	ttg	ctt	cg	aac	gcc	ggg	gac	ctg	gcc	ccc	gct	ggc	ggc	gct	gcc	319
Phe	Leu	Leu	Arg	Asn	Ala	Gly	Asp	Leu	Ala	Pro	Ala	Gly	Gly	Ala	Ala	
		40					45					50				
tcc	gct	agc	acg	gat	gaa	gct	gcc	gac	gcc	gag	agc	ggg	acc	cga	aac	367
Ser	Ala	Ser	Thr	Asp	Glu	Ala	Ala	Asp	Ala	Glu	Ser	Gly	Thr	Arg	Asn	
		55				60					65					
cg	cag	ctg	cag	cag	ctc	atc	tcc	ttt	tcc	atg	gcc	tgg	gcg	aac	cac	415
Arg	Gln	Leu	Gln	Gln	Leu	Ile	Ser	Phe	Ser	Met	Ala	Trp	Ala	Asn	His	
		70			75				80					85		
gtc	ttc	ctc	ggg	tgc	cga	tac	cct	caa	aaa	gtt	atg	gat	aaa	ata	ctt	463
Val	Phe	Leu	Gly	Cys	Arg	Tyr	Pro	Gln	Lys	Val	Met	Asp	Lys	Ile	Leu	
			90					95						100		
agt	atg	gct	gaa	ggc	atc	aa										483
Ser	Met	Ala	Glu	Gly	Ile											
			105													

<210> 613
 <211> 473
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 277..471

<400> 613																
aaact	gtcct	tgccacc	aca	attattt	aat	ggaccca	aaca	gaaag	taacc	ccgga	aatta					60
ggacac	cctca	tcccaaa	aaga	cctttaa	aata	gggga	agtcc	acttg	tgcac	ggctg	ctcct					120
tgctat	agaa	gacct	gggac	agagg	actgc	tgtct	gccct	ctctg	gtcac	cctgc	ctagc					180
tagagg	atct	gtgac	ccag	ccatg	aggac	cctcg	ccatc	cttgct	gcsa	ttctc	ctggt					240

335

ggccctgcag gccagagcca ctccaggcaa gagctg atg agg ttg ctg cag ccc 294
Met Arg Leu Leu Gln Pro
1 5
cgg agc aga ttg cag cgg aca tcc cag aag tgg ttg ttt csc ttg cat 342
Arg Ser Arg Leu Gln Arg Thr Ser Gln Lys Trp Leu Phe Xaa Leu His
10 15 20
ggc gac gaa agc ttg gct cca aag cat cca ggc tca agg aaa aac atg 390
Gly Asp Glu Ser Leu Ala Pro Lys His Pro Gly Ser Arg Lys Asn Met
25 30 35
gac tgc tat tgg cag aat acc agc gtg cat tgc agg aga acg tcg cta 438
Asp Cys Tyr Trp Gln Asn Thr Ser Val His Cys Arg Arg Thr Ser Leu
40 45 50
tgg aac ctg cat cta cca ggg aag act ctg ggc at 473
Trp Asn Leu His Leu Pro Gly Lys Thr Leu Gly
55 60 65

<210> 614

<211> 522

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 279..521

<400> 614

attctyaatt gsgccmagct tccggccttg ccacccgcca gcaggtctca cccggttcgc 60
atcagtcctc ggcccccgcc gccctcagc agtagcaaat tatcgcatcc ctgagcttag 120
gatccccctt tagtttcaca tatgccccca ccccggttcc cactttgatg gcggcttcca 180
aggagcatcc cttagtaaga accgtcgcaa atatcgtgat ttggatccac caaggcttgg 240
ggcgagccgt cgaatgtgtg tggtagggga gcgggctg atg tta gca cgt ggt ctt 296
Met Leu Ala Arg Gly Leu
1 5
gtt arg scy tgr aat kgg ggr aat agg ags cck gtg ttc acg att agt 344
Val Xaa Xaa Xaa Asn Xaa Gly Asn Arg Xaa Pro Val Phe Thr Ile Ser
10 15 20
ttc cac cac cta cca act tgg aac ctt tac atc tct cca agc ctg agt 392
Phe His His Leu Pro Thr Trp Asn Leu Tyr Ile Ser Pro Ser Leu Ser
25 30 35
ttc ctt atc tgt gag aca gct gcg aat acc ggt ttc cag aga gta gtt 440
Phe Leu Ile Cys Glu Thr Ala Ala Asn Thr Gly Phe Gln Arg Val Val
40 45 50
aag aac cct tgt gaa cac ctt ctt cag ctt cct tcg acg caa aac aga 488
Lys Asn Pro Cys Glu His Leu Leu Gln Leu Pro Ser Thr Gln Asn Arg
55 60 65 70
ctt ttt cat tgg agg aga aga agg gat ggc aga g 522
Leu Phe His Trp Arg Arg Arg Arg Asp Gly Arg
75 80

<210> 615

<211> 322

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 171..320

<400> 615

agtgttccg ccgcggcgac cccactccgg cggcgcgtcc cccgagcttg gtacggctca 60
gccgtctcc ccgaagccg cgcgcccgcg cccgcgcccc tcagtcgggtg gagcccgag 120
cccccttgt ggccccgggc agctccccgc ccgtcgggcc cgcgcccgcc atg gtc 176
Met Val


```

                                1
cgt ccg cgc cgt gcc ccg tac cgc tcc ggc gcc ggg ggc ccc ctm ggg      224
Arg Pro Arg Arg Ala Pro Tyr Arg Ser Gly Ala Gly Gly Pro Leu Gly
      5              10              15
ggg cgc ggc cgy cct ccg cgg ccc ctc gtg gtg cgc gcc gtc cgc tcg      272
Gly Arg Gly Arg Pro Pro Arg Pro Leu Val Val Arg Ala Val Arg Ser
      20              25              30
cgc tcc tgg cct gcc agc ccy ccg agg ccc gca gcc tyc gcg gat ccg      320
Arg Ser Trp Pro Ala Ser Pro Pro Arg Pro Ala Ala Xaa Ala Asp Pro
35              40              45              50
gg

```

<210> 616
 <211> 524
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 269..523

```

<400> 616
attttataact aatggagaga gagagaaaga catagaagtc tcaactgactt gcacagttta      60
cttcccactg actagatctt ggataaggac ttttttgtct tattcataaa aggcgattgt      120
gttcaattat gttatgctca gaggtggaat atataatact gacttctatt gcttaaaaaa      180
ggtgattatg aatggaggaa gaagtctgtg ggacaggaaa tttaaaataa aatttttagag      240
ttagatagat tatgagaaga aaaaatca atg cca agt ttt gtt gac cgc atc      292
                        Met Pro Ser Phe Val Asp Arg Ile
                        1              5
ttt ggt ggt gaa cta act agt atg atc atg tat gat caa tgc aga act      340
Phe Gly Gly Glu Leu Thr Ser Met Ile Met Tyr Asp Gln Cys Arg Thr
      10              15              20
gtc tcc ttg gtt cat gaa tct ttc ctt gat ttg tcc ctc cca gtt tta      388
Val Ser Leu Val His Glu Ser Phe Leu Asp Leu Ser Leu Pro Val Leu
25              30              35              40
gat gat cag agt ggt aag aaa agt gta aat gat aaa aat ctg aaa aag      436
Asp Asp Gln Ser Gly Lys Lys Ser Val Asn Asp Lys Asn Leu Lys Lys
      45              50              55
aca gtg gag gat gaa gat caa gat agt gag gaa gaa aaa gat aac gac      484
Thr Val Glu Asp Glu Asp Gln Asp Ser Glu Glu Glu Lys Asp Asn Asp
      60              65              70
agt tac ata aaa gag aga agt gat att cct tct gga aca a      524
Ser Tyr Ile Lys Glu Arg Ser Asp Ile Pro Ser Gly Thr
      75              80              85

```

<210> 617
 <211> 274
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 75..272

```

<400> 617
atacacttct tggctgtgtg cgctcagcag gacgtgggag gctccggctt caagacactc      60
atccaagagg aagg atg gcc agt atc ttt tct aag ttg cta act ggc cgc      110
                        Met Ala Ser Ile Phe Ser Lys Leu Leu Thr Gly Arg
                        1              5              10
aat gct tct ctg ctg ttt gct acc atg ggc acc agt gtc ctg acc acc      158
Asn Ala Ser Leu Leu Phe Ala Thr Met Gly Thr Ser Val Leu Thr Thr
      15              20              25
ggg tac ctg ctg aac cgg cag aaa gtg tgt gcc gag gtc cgg gag cag      206

```

337

Gly	Tyr	Leu	Leu	Asn	Arg	Gln	Lys	Val	Cys	Ala	Glu	Val	Arg	Glu	Gln	
30						35					40					
cct	agg	cta	ttt	cct	cca	agc	gca	gac	tac	cca	gac	ctg	cgc	aas	aca	254
Pro	Arg	Leu	Phe	Pro	Pro	Ser	Ala	Asp	Tyr	Pro	Asp	Leu	Arg	Xaa	Thr	
45					50					55					60	
aca	act	gca	tggt	ccg	agt	gc										274
Thr	Thr	Ala	Trp	Pro	Ser											
				65												

<210> 618
 <211> 243
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 21..242

<400> 618																
agggaagagg	ctgtcgggaa	atg	gcg	gcc	gcg	gcc	ggg	ctg	ggg	ctt	cag	cgg				53
		Met	Ala	Ala	Ala	Ala	Gly	Leu	Gly	Leu	Gln	Arg				
		1				5				10						
gag	gca	gca	gag	ggg	aag	tgg	tca	gcg	tgg	cga	atg	acg	gaa	gaa	act	101
Glu	Ala	Ala	Glu	Gly	Lys	Trp	Ser	Ala	Trp	Arg	Met	Thr	Glu	Glu	Thr	
		15				20				25						
cgc	att	gtc	tac	tgg	atc	aag	gac	aga	cag	ctc	acc	aac	cgt	gac	agc	149
Arg	Ile	Val	Tyr	Trp	Ile	Lys	Asp	Arg	Gln	Leu	Thr	Asn	Arg	Asp	Ser	
		30				35				40						
acc	ata	ctg	gaa	ctt	caa	aaa	gtt	ctg	aaa	aca	tgt	tgg	ttc	gtt	tgt	197
Thr	Ile	Leu	Glu	Leu	Gln	Lys	Val	Leu	Lys	Thr	Cys	Trp	Phe	Val	Cys	
		45				50				55						
tgg	agt	att	ttt	caa	ctt	aaa	ttc	ctt	ttc	ctt	ttt	ttt	ttt	ttt	k	243
Trp	Ser	Ile	Phe	Gln	Leu	Lys	Phe	Leu	Phe	Leu	Phe	Phe	Phe	Phe		
60					65					70						

<210> 619
 <211> 499
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 295..498

<400> 619																
aggaagttga	tcaatcccga	ggctgctgag	agacggtggc	gcgattggga	cagtcgccag											60
ggatggctga	gcgtgaagat	gcagcgggtg	tccgggctgc	tctcctggac	gctgagcaga											120
gtcctgtggc	tctccggcct	ctctgagccg	ggagctgccc	ggcagccccg	gatcatggaa											180
gagaaagcct	agaggtttat	gatttgatta	gaactatccg	ggacccagaa	aagccaata											240
ctttagaaga	actggaagtg	gtctcggaaa	gttgtgtgga	agttcaggag	ataa	atg										297
						Met										
						1										
aag	aag	aat	atc	tgg	tta	tta	tca	ggt	tca	cgc	caa	cag	tac	ctc	att	345
Lys	Lys	Asn	Ile	Trp	Leu	Leu	Ser	Gly	Ser	Arg	Gln	Gln	Tyr	Leu	Ile	
		5						10					15			
gct	ctt	tgg	cga	ctc	tta	ttg	ttg	gaa	atc	tac	att	tct	gaa	gga	acc	393
Ala	Leu	Trp	Arg	Leu	Leu	Leu	Leu	Glu	Ile	Tyr	Ile	Ser	Glu	Gly	Thr	
		20				25						30				
cac	tca	aca	gaa	gaa	gac	atc	aat	aag	cag	ata	aat	gac	aaa	gag	cga	441
His	Ser	Thr	Glu	Glu	Asp	Ile	Asn	Lys	Gln	Ile	Asn	Asp	Lys	Glu	Arg	
		35				40					45					
gtg	gca	gct	gca	atg	gaa	aac	ccc	aac	tta	cgg	gaa	att	gtg	gac	agt	489
Val	Ala	Ala	Ala	Met	Glu	Asn	Pro	Asn	Leu	Arg	Glu	Ile	Val	Asp	Ser	

338

50 55 60 65 499
 gtg tcc ttg a
 Val Ser Leu

<210> 620
 <211> 771
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 613..771

<400> 620
 cagggttcagg gcacagaacc caggcttgta ccatgggtggg gggagaaaat gaccactggc 60
 caagaggact gctgacctgt gcaccaggct agtacttatg actacaaatt cttactgctt 120
 ctctaataca ctctgagggg agagggcatc tgatcattac aaaagggagg gcttataagt 180
 gatctcccaa gaaggcagtg atctgctagt gcctttggct ctgtacctct gctgggcatc 240
 tctccaaggt ctaaggtaac atattaaatg tttttgtcag ctaatgcagg ctcagtgact 300
 ttaagtctgt aagttacca ggaagaaggg ttataggaaa aataactcag taagtttaaa 360
 accaaacaca tttccattta gtgacaggaa tttaagcaag gacctgaagt agaatacaact 420
 gattcacaca gtagtaaata caaagtagaa caatgatctt ggcttcgctg tctgggttcag 480
 tggctgctgt gaatgcaata cacaagttaa gtcacactgc agactgtttt ctagctgtgg 540
 ccgctggatg ccacttctag catagtagaa ctatgttagg aggaatggga aaagtgagca 600
 ccacttctca cc atg ttc ccc cct cct gct gcc agt ctc tgc tcc cat gtt 651
 Met Phe Pro Pro Pro Ala Ala Ser Leu Cys Ser His Val
 1 5 10
 gga tgc agc aga gat cac cca cca gtt ggc cca gga cag acc aat agg 699
 Gly Cys Ser Arg Asp His Pro Pro Val Gly Pro Gly Gln Thr Asn Arg
 15 20 25
 aag ggt cca atc act cta act aca gcc gaa ctc acc tcc aca aca gtc 747
 Lys Gly Pro Ile Thr Leu Thr Thr Ala Glu Leu Thr Ser Thr Thr Val
 30 35 40 45
 tct gtg gct cta gcc tgg act cct 771
 Ser Val Ala Leu Ala Trp Thr Pro
 50

<210> 621
 <211> 658
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 440..658

<400> 621
 gctttgtaat tgggtgcagct ctgagaacac acgcatgcat tccaggcggt gctgctgggg 60
 ctcggcgggg ccgtgtcttg tctgcaaggc gcgcgggctc gtgggggggtt ggctggggag 120
 cccacgctgc ctggcgactc gggccaccga atgtgagacc gagtcccttt atgtcaccag 180
 cgcacacgct gatttgaacc ctgcttcgac gtgtgtgtca tggcttaaaa atagctgcta 240
 atctgtcaac ctgtcttggg cagaaacagc ggcggcgaca gagcaggagc gtcattggccg 300
 tggcgctgtc tgcgcsggcg atccgccttt cggactgagg cccagcgag cgcttgcaaa 360
 gagcagcagc tacctggcaa ctgaacccat catcaccaca gccactcctg cagctgccac 420
 ggtttctgcc acctctaag atg tgc cct ggt aac tgg ctt tgg gct tct atg 472
 Met Cys Pro Gly Asn Trp Leu Trp Ala Ser Met
 1 5 10
 act ttt atg gcc cgc ttc tcc cgg agt agc tca agg tct cct gtt cga 520
 Thr Phe Met Ala Arg Phe Ser Arg Ser Ser Ser Arg Ser Pro Val Arg
 15 20 25
 act cga ggg acc ctg gag gag atg cca acc gtt caa cat cct ttc ctc 568
 Thr Arg Gly Thr Leu Glu Glu Met Pro Thr Val Gln His Pro Phe Leu
 30 35 40

340

aca gcc atc atg ggg cta gat gaa cag gct aga acc ctg gaa agg gaa 219
 Thr Ala Ile Met Gly Leu Asp Glu Gln Ala Arg Thr Leu Glu Arg Glu
 10 15 20 25
 gag aag aaa gaa aac agt gga gca atc aga tca gac aga tgt gta aag 267
 Glu Lys Lys Glu Asn Ser Gly Ala Ile Arg Ser Asp Arg Cys Val Lys
 30 35 40
 gca tca ttc cag cag gga ctg gca aac aat att tcc ttt tat cta tgc 315
 Ala Ser Phe Gln Gln Gly Leu Ala Asn Asn Ile Ser Phe Tyr Leu Cys
 45 50 55
 cca cct ggc tac agt cca cgt atg ctc aga ttc tct tct gcc tct gcc 363
 Pro Pro Gly Tyr Ser Pro Arg Met Leu Arg Phe Ser Ser Ala Ser Ala
 60 65 70
 gcc ctg aga cag caa gac caa ccc ctt ctc ctc ctc ttc ctc ctc ctc 411
 Ala Leu Arg Gln Gln Asp Gln Pro Leu Leu Leu Leu Phe Leu Leu Leu
 75 80 85
 agc cta ctc aac gtg aag agg atg aga atg aag acc ttt atg atg acc 459
 Ser Leu Leu Asn Val Lys Arg Met Arg Met Lys Thr Phe Met Met Thr
 90 95 100 105
 cgc atc cac tta atg aat agt a 481
 Arg Ile His Leu Met Asn Ser
 110

<210> 624
 <211> 456
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 295..456

<400> 624
 aggcggtgct gcctgtgtcg tcgtccctgt ttgttcttct tcaaacctca ggatttctcc 60
 gtgtttccca gaccgtaacg ttccaaaggt agaaacacag atgacagatt tttggtcttt 120
 tcgtcatttc ctatatatgt gttcaatggt ttggaatagt tcaccagtga agtcatctca 180
 accagcacag ggccttgact cctaacaaca atcacacgaa cttggaagag gatccttccc 240
 cagctgaaac ttcacttgag accttctcat tggccatcat cgacatctgg attc atg 297
 Met
 1
 ata cag aga aac ttg agt aat gat gac aat agc agt cca gcc tgc atg 345
 Ile Gln Arg Asn Leu Ser Asn Asp Asp Asn Ser Ser Pro Ala Cys Met
 5 10 15
 gat gct gac atg gat gct gtc tcc cat tct tac cac tgg atg tca aga 393
 Asp Ala Asp Met Asp Ala Val Ser His Ser Tyr His Trp Met Ser Arg
 20 25 30
 aat gct aga tgc gat cca gcc agt ctg tac tcg act ctc agt tcc atc 441
 Asn Ala Arg Cys Asp Pro Ala Ser Leu Tyr Ser Thr Leu Ser Ser Ile
 35 40 45
 ttc tca act cac agt 456
 Phe Ser Thr His Ser
 50

<210> 625
 <211> 551
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 137..550

<400> 625
 aaaaaaac agctgagtct ttgtgccagg aagactgcgt gcagaaggtg actgtctcag 60

341

```

tggagctggg tcatctcagg ccttggtctc ttgaactttt ggccgccatg tgcttcccga 120
agtcctctct gatgac atg aag aag ctg aag gcc cga atg cac cag gcc ata 172
      Met Lys Lys Leu Lys Ala Arg Met His Gln Ala Ile
      1          5          10
gaa aga ttt tat gat aaa atg caa aat gca gaa tca gga cgt gga cag 220
Glu Arg Phe Tyr Asp Lys Met Gln Asn Ala Glu Ser Gly Arg Gly Gln
      15          20          25
gtg atg tcg agc ctg gca gag ctg gag gac gac ttc aaa gag ggc tac 268
Val Met Ser Ser Leu Ala Glu Leu Glu Asp Asp Phe Lys Glu Gly Tyr
      30          35          40
ctg gag aca gtg gcg gct tat tat gag gag cag cac cca gag ctc act 316
Leu Glu Thr Val Ala Ala Tyr Tyr Glu Glu Gln His Pro Glu Leu Thr
      45          50          55          60
cct cta ctt gaa aaa gaa aga gat gga tta cgg tgc cga ggc aac aga 364
Pro Leu Leu Glu Lys Glu Arg Asp Gly Leu Arg Cys Arg Gly Asn Arg
      65          70          75
tcc cct gtc ccg gat gtt gag gat ccc gca acc gag gag cct ggg gag 412
Ser Pro Val Pro Asp Val Glu Asp Pro Ala Thr Glu Glu Pro Gly Glu
      80          85          90
agc ttt tgt gac aag gtc atg aga tgg ttc cag gcc atg ctg cag cgg 460
Ser Phe Cys Asp Lys Val Met Arg Trp Phe Gln Ala Met Leu Gln Arg
      95          100          105
ctg cag acc tgg tgg cac ggg gtt ctg gcc tgg gtg aag gag aag gtg 508
Leu Gln Thr Trp Trp His Gly Val Leu Ala Trp Val Lys Glu Lys Val
      110          115          120
gtg gcc ctg gtc cat gca gtg cag gcc ctc tgg aaa cag ttc c 551
Val Ala Leu Val His Ala Val Gln Ala Leu Trp Lys Gln Phe
      125          130          135

```

<210> 626

<211> 494

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 283..492

<400> 626

```

gacttttgca aggcattggcg gggacactgt gaatgtcagc ccagaagggtg atcagagcct 60
gttaattaaa atggaaagaa gacagaaggg aaggtagaca tcagggttctc cctggagact 120
tttcgttttc atttacgctg cggaaactga cgtttttgcc taacaccccca tgtaatgtaa 180
acgtataggc ttgagtagct gtccggccgc atgtgtagtg aaccctaaag ctttcctaata 240
tgtagtttagc atcgtcccta agcgggaacga ttttcctgta ac atg att tgt act 294
      Met Ile Cys Thr
      1

```

```

ttt cta cga gcc gta cag tat acg gag aag ctg cac agg tcc tcg gca 342
Phe Leu Arg Ala Val Gln Tyr Thr Glu Lys Leu His Arg Ser Ser Ala
      5          10          15          20

```

```

aag cga ttg ctt ttg cca tac atc gtg ctt aac aaa gcg tgc ttg aag 390
Lys Arg Leu Leu Leu Pro Tyr Ile Val Leu Asn Lys Ala Cys Leu Lys
      25          30          35

```

```

act gag ccc agt ttg aga tgt ggg ctt caa tat caa aag aaa acg ctg 438
Thr Glu Pro Ser Leu Arg Cys Gly Leu Gln Tyr Gln Lys Lys Thr Leu
      40          45          50

```

```

cga cct aga tgt att ctt gga gtc acc cag aaa acc atc tgg acg cag 486
Arg Pro Arg Cys Ile Leu Gly Val Thr Gln Lys Thr Ile Trp Thr Gln
      55          60          65

```

```

gga ccg ag 494
Gly Pro
      70

```

<210> 627

<211> 585
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 128..583

<400> 627
 attttccggg aagatggcgg cgcacaagtc aggtccggca catgtttccg cggagcccca 60
 gcaatgacgg atgatatcac ctcttcttct ctggtgagag tctgaggata gagacttttt 120
 tctcacc atg aat gtc acc cca gag gtc aag agt cgt ggg atg aag ttt 169
 Met Asn Val Thr Pro Glu Val Lys Ser Arg Gly Met Lys Phe
 1 5 10
 gct gag gag cag ctg cta aag cat gga tgg act caa ggc aaa ggc ctc 217
 Ala Glu Glu Gln Leu Leu Lys His Gly Trp Thr Gln Gly Lys Gly Leu
 15 20 25 30
 ggc cgg aag gag aat ggt atc act cag gct ctc agg gtg aca ctg aag 265
 Gly Arg Lys Glu Asn Gly Ile Thr Gln Ala Leu Arg Val Thr Leu Lys
 35 40 45
 caa gac act cat ggg gta gga cat gac cct gcc aag gag ttc aca aac 313
 Gln Asp Thr His Gly Val Gly His Asp Pro Ala Lys Glu Phe Thr Asn
 50 55 60
 cac tgg tgg aat gag ctc ttc aac aag act gcg gcc aac ttg gta gtg 361
 His Trp Trp Asn Glu Leu Phe Asn Lys Thr Ala Ala Asn Leu Val Val
 65 70 75
 gaa act ggg cag gat gga gta cag ata agg agc ctt tct aag gag acc 409
 Glu Thr Gly Gln Asp Gly Val Gln Ile Arg Ser Leu Ser Lys Glu Thr
 80 85 90
 acc cgt tat aat cat ccc aag ccc aac ttg ctg tat cag aag ttt gtg 457
 Thr Arg Tyr Asn His Pro Lys Pro Asn Leu Leu Tyr Gln Lys Phe Val
 95 100 105 110
 aag atg gct aca ttg act tca ggt gga gag aag cca aac aaa gac ttg 505
 Lys Met Ala Thr Leu Thr Ser Gly Gly Glu Lys Pro Asn Lys Asp Leu
 115 120 125
 gag agc tgc agt gat gac gac aac cag ggg tca agt ccc caa aga ttc 553
 Glu Ser Cys Ser Asp Asp Asp Asn Gln Gly Ser Ser Pro Gln Arg Phe
 130 135 140
 tta ctg atg aga tgc tgc tcc aag cct gtg ar 585
 Leu Leu Met Arg Cys Cys Ser Lys Pro Val
 145 150

<210> 628
 <211> 560
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 201..560

<400> 628
 atcytawtcg aaaagctctt tccttcttcc tcttggtcct cctcctgcct ctcttcgctt 60
 cgctgcaaaa cgcggtgggg gctgctcggc ggtcaggagc agcaagagac agagcgacat 120
 gagagattgg accgcgggct gcactggaga atttactggt aggataattc atccctaaag 180
 agattgaagt gagcttcaga atg gca aaa gag gag ccc cag agt atc tca agg 233
 Met Ala Lys Glu Glu Pro Gln Ser Ile Ser Arg
 1 5 10
 gac ttg cag gaa ctg cag aag aag ctg tct ctg ctg ata gac tcc ttc 281
 Asp Leu Gln Glu Leu Gln Lys Lys Leu Ser Leu Leu Ile Asp Ser Phe
 15 20 25
 cag aat aac tca aag gtg gtg gcc ttt atg aag tct cca gtg ggt cag 329
 Gln Asn Asn Ser Lys Val Val Ala Phe Met Lys Ser Pro Val Gly Gln

343

30	35	40	
tac ttg gac agc cat ccg ttt ctg gcc ttc acc ttg ctg gtg ttc att			377
Tyr Leu Asp Ser His Pro Phe Leu Ala Phe Thr Leu Leu Val Phe Ile			
45	50	55	
gtc atg tcg gcc gtt cct gtt gga ttc ttc ctg ctc atc gtg gtg ctt			425
Val Met Ser Ala Val Pro Val Gly Phe Phe Leu Leu Ile Val Val Leu			
60	65	70	75
acc acc ctg gct gct ctg ctg ggg gtc ata ata ttg gaa gga ttg gtc			473
Thr Thr Leu Ala Ala Leu Leu Gly Val Ile Ile Leu Glu Gly Leu Val			
80	85	90	
atc tct gtg ggt ggc ttc tca ctg ctc tgc atc ctc tgt ggk ttg agc			521
Ile Ser Val Gly Gly Phe Ser Leu Leu Cys Ile Leu Cys Gly Leu Ser			
95	100	105	
ttc gta tca ctc gcc atg tcg ggg atg atg ata gca tct			560
Phe Val Ser Leu Ala Met Ser Gly Met Met Ile Ala Ser			
110	115	120	

<210> 629

<211> 728

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 420..728

<400> 629

tttttggagg gctcttagca acggccctgg ttgagccccc tcagccatga gaaaatcaaa	60		
tcaatgtgcc attcttccag gcgcgaggca gcagcggctg cagttcaaca tgaaaggagg	120		
cttcctccct gctgtataat tacctgctct tcccgatctc atcgtttctg cctttgcaaa	180		
gtgctactga gaagsgggaa gaaacgtccg ccacccatcc cccttgctgc ctgggggttc	240		
agacttgatt agatggctaa caggggccc agctatggct taagccgaga ggtgcaggag	300		
aagatcgagc agaagtatga tgcggacctg gagaacaagc tgggtggactg gatcatcctg	360		
cagtgcgccg aggacataga gcacccgccc cccggcaggc ccattttcag aaatggtta	419		
atg gac ggg acg gtc ctg tgc aag ctg ata aat agt tta tac cca cca	467		
Met Asp Gly Thr Val Leu Cys Lys Leu Ile Asn Ser Leu Tyr Pro Pro			
1	5	10	15
gga caa gag ccc ata ccc aag atc tca gag tca aag atg gct ttt aag	515		
Gly Gln Glu Pro Ile Pro Lys Ile Ser Glu Ser Lys Met Ala Phe Lys			
20	25	30	
cag atg gag caa atc tcc cag ttc cta aaa gct gcg gag acc tat ggt	563		
Gln Met Glu Gln Ile Ser Gln Phe Leu Lys Ala Ala Glu Thr Tyr Gly			
35	40	45	
gtc aga acc acc gac atc ttt cag acg gtg gat cta tgg gaa ggg aag	611		
Val Arg Thr Thr Asp Ile Phe Gln Thr Val Asp Leu Trp Glu Gly Lys			
50	55	60	
gac atg gca gct gtg cag agg acc ctg atg gct tta ggc agc gtt gca	659		
Asp Met Ala Ala Val Gln Arg Thr Leu Met Ala Leu Gly Ser Val Ala			
65	70	75	80
gtc acc aag gat gat ggc tgc tat cgg gga gag cca tcc tgg ttt cac	707		
Val Thr Lys Asp Asp Gly Cys Tyr Arg Gly Glu Pro Ser Trp Phe His			
85	90	95	
agg aaa gcc cag cag aat cgg	728		
Arg Lys Ala Gln Gln Asn Arg			
100			

<210> 630

<211> 496

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 261..494

<400> 630

```

ttcgcggtta gggattccat agttcacggg ttacggcaca cacgcgttac gtatatatcg      60
aggttaatct gcgcaastct tcacttttgt ctgaatgtat tttagcgggtg aattactcct      120
tctggctaag agcttcacat cttcagtggtt aatgggtgtg gttttcgcac gtcttgacaca      180
catttctaga tcttgcaactg tcgaaaagtc agctccgaaa ttgctgcaat gataagtttg      240
ctgaactgca tctctgcccc atg cca ctt cct cgc aaa gac aac cca cag tgt      293
                               Met Pro Leu Pro Arg Lys Asp Asn Pro Gln Cys
                               1         5         10
agt cca ctg ctg cct cct ttg gcc tct gtt ggt aag aaa atg gct aag      341
Ser Pro Leu Leu Pro Pro Leu Ala Ser Val Gly Lys Lys Met Ala Lys
                               15         20         25
aaa atc gct gct gct cct cca cct cca cct cct cct cca tca gaa cgg      389
Lys Ile Ala Ala Ala Pro Pro Pro Pro Pro Pro Pro Pro Ser Glu Arg
                               30         35         40
cga ggc aaa cgc gca gaa cag aaa ggg gct gtg ggg cag cgg cca gat      437
Arg Gly Lys Arg Ala Glu Lys Gly Ala Val Gly Gln Arg Pro Asp
                               45         50         55
tgg att ttc tct ccg cag ccc caa rgg atg ttt cta atg acc caa gga      485
Trp Ile Phe Ser Pro Gln Pro Gln Xaa Met Phe Leu Met Thr Gln Gly
60                               65                               70                               75
aaw tta gtg gt      496
Xaa Leu Val

```

<210> 631

<211> 500

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 91..498

<400> 631

```

wgaacatcct aatcgaaaaa agattctaag acaagacttc taatacggca tctcagttta      60
ttgttctttg gggtttgttt tcagatcctc atg tca gaa cca ggc aaa ctc agc      114
                               Met Ser Glu Pro Gly Lys Leu Ser
                               1         5
cag aaa atc aaa gtt tgg ctt cag gag tac tgg aac atc aca gat ctc      162
Gln Lys Ile Lys Val Trp Leu Gln Glu Tyr Trp Asn Ile Thr Asp Leu
                               10         15         20
gtg gcc att tcc aca ttc atg att gga gca att ctt cgc cta cag aac      210
Val Ala Ile Ser Thr Phe Met Ile Gly Ala Ile Leu Arg Leu Gln Asn
25                               30         35         40
cag ccc tac atg ggc tat ggc cgg gtg atc tac tgt gtg gat atc atc      258
Gln Pro Tyr Met Gly Tyr Gly Arg Val Ile Tyr Cys Val Asp Ile Ile
                               45         50         55
ttc tgg tac atc cgt gtc ctg gac atc ttt ggt gtc aac aag tat ctg      306
Phe Trp Tyr Ile Arg Val Leu Asp Ile Phe Gly Val Asn Lys Tyr Leu
60                               65                               70
ggg cca tac gtg atg atg att gga aag atg atg atc gac atg ctg tac      354
Gly Pro Tyr Val Met Met Ile Gly Lys Met Met Ile Asp Met Leu Tyr
75                               80         85
ttt gtg gtc atc atg ctg gtc gtg ctc atg agt ttc gga gta gcc cgt      402
Phe Val Val Ile Met Leu Val Val Leu Met Ser Phe Gly Val Ala Arg
90                               95         100
caa gcc att ctg cat cca gag gag aag ccc tct tgg aaa ctg gcc cga      450
Gln Ala Ile Leu His Pro Glu Glu Lys Pro Ser Trp Lys Leu Ala Arg
105                               110         115         120
aac atc ttc tac atg ccc tac tgg atg atc tat gga gag gtg ttg cag      498
Asn Ile Phe Tyr Met Pro Tyr Trp Met Ile Tyr Gly Glu Val Leu Gln
125                               130         135

```

ac

500

<210> 632
 <211> 524
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 170..523

<400> 632
 tttttccggg cttactcaag ttgcggcctt cctcgcgtca cagccgggat gaagccgac 60
 ctactgcagg gccatgagcg gtccattacg cagattaagt ataaccgcga aggagacctc 120
 ctctttactg tggccaagga ccctatcgctc aatgtatggg actctgtga atg gtg aga 178
 Met Val Arg
 1
 ggc tgg gca cct aca tgg gcc ata ccg gag ctg tgt ggt gtg tgg acg 226
 Gly Trp Ala Pro Thr Trp Ala Ile Pro Glu Leu Cys Gly Val Trp Thr
 5 10 15
 ctg act ggg aca cca agc atg tcc tca ctg gct cag ctg aca aca gct 274
 Leu Thr Gly Thr Pro Ser Met Ser Ser Leu Ala Gln Leu Thr Thr Ala
 20 25 30 35
 gtc gtc tct ggg act gtg aaa cag tct gga gag gtg ttg gtg aat gtt 322
 Val Val Ser Gly Thr Val Lys Gln Ser Gly Glu Val Leu Val Asn Val
 40 45 50
 aag gag cac tcc cgg cag atc aac gac atc cag tta tcc agg gac atg 370
 Lys Glu His Ser Arg Gln Ile Asn Asp Ile Gln Leu Ser Arg Asp Met
 55 60 65
 acc atg ttt gtg acc gcg tcc aag gac aac aca gcc aag ctt ttt gac 418
 Thr Met Phe Val Thr Ala Ser Lys Asp Asn Thr Ala Lys Leu Phe Asp
 70 75 80
 tcc aca act ctt gaa cat cag aag act ttc cgg aca gaa cgt cct gtc 466
 Ser Thr Thr Leu Glu His Gln Lys Thr Phe Arg Thr Glu Arg Pro Val
 85 90 95
 aac tca gct gcc ctg tcc ccc aac tat gac cat gtg gtc ctg ggc ggt 514
 Asn Ser Ala Ala Leu Ser Pro Asn Tyr Asp His Val Val Leu Gly Gly
 100 105 110 115
 ggt cag gaa g 524
 Gly Gln Glu

<210> 633
 <211> 508
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 141..506

<400> 633
 attggcttct ctcccagttt ccaccacact ccccatctgg gtaacctaag caggttactc 60
 gccctgtgga ggaggttctc ttctctgtac aatggaaatg gttatccaca cctatctctc 120
 agggaggctc tgatatcacc atg aga ccc agc tgt gct cct gga tgg ttt tac 173
 Met Arg Pro Ser Cys Ala Pro Gly Trp Phe Tyr
 1 5 10
 cac aag tcc aat tgc tat ggt tac ttc agg aag ctg agg aac tgg tct 221
 His Lys Ser Asn Cys Tyr Gly Tyr Phe Arg Lys Leu Arg Asn Trp Ser
 15 20 25
 gat gcc gag ctg gag tgt cag tct tac gga aac gga gcc cac ctg gca 269
 Asp Ala Glu Leu Glu Cys Gln Ser Tyr Gly Asn Gly Ala His Leu Ala
 30 35 40
 tct atc ctg agt tta aag gaa gcc agc acc ata gca gag tac ata agt 317

346

Ser	Ile	Leu	Ser	Leu	Lys	Glu	Ala	Ser	Thr	Ile	Ala	Glu	Tyr	Ile	Ser		
45						50					55						
ggc	tat	cag	aga	agc	cag	ccg	ata	tgg	att	ggc	ctg	cac	gac	cca	cag		365
Gly	Tyr	Gln	Arg	Ser	Gln	Pro	Ile	Trp	Ile	Gly	Leu	His	Asp	Pro	Gln		
60					65					70					75		
aag	agg	cag	cag	tgg	cag	tgg	att	gat	ggg	gcc	atg	tat	ctg	tac	aga		413
Lys	Arg	Gln	Gln	Trp	Gln	Trp	Ile	Asp	Gly	Ala	Met	Tyr	Leu	Tyr	Arg		
				80					85					90			
tcc	tgg	tct	ggc	aag	tcc	atg	ggt	ggg	aac	aag	cac	tgt	gct	gag	atg		461
Ser	Trp	Ser	Gly	Lys	Ser	Met	Gly	Gly	Asn	Lys	His	Cys	Ala	Glu	Met		
			95					100					105				
agc	tcc	aat	aac	aac	ttt	tta	act	tgg	agc	agc	aac	gaa	tgc	aac	aa		508
Ser	Ser	Asn	Asn	Asn	Phe	Leu	Thr	Trp	Ser	Ser	Asn	Glu	Cys	Asn			
		110					115					120					

<210> 634

<211> 464

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 92..412

<220>

<221> sig_peptide

<222> 92..160

<223> Von Heijne matrix

score 12

seq RLLLLSLVSGALG/LG

<220>

<221> misc_feature

<222> 325

<223> n=a, g, c or t

<400> 634

tgctgccttc	tctgatggct	tctccagcgg	cctcagtacc	accttagcgg	tcttctgcc		60
tgagctgccc	actaactgg	gtgactttgc	c	atg	ctg	ctc	112
			Met	Leu	Leu	Gln	
							-20
tcc	ttt	cgg	cgg	ctg	ctg	ctg	160
Ser	Phe	Arg	Arg	Leu	Leu	Leu	
	-15			-10		-5	
ttg	ggg	ggt	gca	gtc	ctg	ggg	208
Leu	Gly	Gly	Ala	Val	Leu	Gly	
1		5			10	15	
ctc	act	ccc	tgg	gtg	ttt	ggg	256
Leu	Thr	Pro	Trp	Val	Phe	Gly	
		20			25	30	
gcc	ctt	gtg	gac	atg	cta	cca	304
Ala	Leu	Val	Asp	Met	Leu	Pro	
	35				40	45	
cct	acg	ccc	cat	gtg	ctc	ctn	352
Pro	Thr	Pro	His	Val	Leu	Leu	
	50				55	60	
ggc	ctc	atg	ctt	gcc	ata	acc	400
Gly	Leu	Met	Leu	Ala	Ile	Thr	
65				70		75	
acc	act	gag	ggc	tgatggggcc	aattggaaag	gggtcgggtt	452
Thr	Thr	Glu	Gly				
cccccaacc	ac						464

```
<210> 635
<211> 506
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 193..468
```

```
<220>
<221> sig_peptide
<222> 193..264
<223> Von Heijne matrix
      score 11.6999998092651
      seq FVLGLGLTPPTLA/QD
```

<400>	635	
acacacactc	acacaaggac	gcsaacccca cctagatgca aagcaggatt caaaagaaca 60
tctttgcgtt	ttctaccggc	tccccatcat cgtactaggg aggaagaagc gggtgagaaa 120
caaaacttgt	ttccattgtc	ctgcccggtt ctgcggactt gttctgaggc cgaggagcct 180
gtgttggaag	ag atg gtg atg ggc ctg ggc gtt ttg ttg ttg gtc ttc gtg 231	
	Met Val Met Gly Leu Gly Val Leu Leu Val Phe Val	
	-20	-15
ctg ggt ctg ggt ctg acc cca ccg acc ctg gct cag gat aac tcc agg 279		
Leu Gly Leu Gly Leu Thr Pro Pro Thr Leu Ala Gln Asp Asn Ser Arg		
	-10	-5 1 5
tac aca cac ttc ctg acc cag cac tat gat gcc aaa cca cag ggc cgg 327		
Tyr Thr His Phe Leu Thr Gln His Tyr Asp Ala Lys Pro Gln Gly Arg		
	10	15 20
gat gac aga tac tgt gaa agc atc atg agg aga cgg ggc ctg acc tca 375		
Asp Asp Arg Tyr Cys Glu Ser Ile Met Arg Arg Arg Gly Leu Thr Ser		
	25	30 35
ccc tgc aaa gac atc aac aca ttt att cat ggc aac aag cgc asa tca 423		
Pro Cys Lys Asp Ile Asn Thr Phe Ile His Gly Asn Lys Arg Xaa Ser		
	40	45 50
agg cca tct gtg aaa aca aga atg gaa acc ctc aca gag aaa acc 468		
Arg Pro Ser Val Lys Thr Arg Met Glu Thr Leu Thr Glu Lys Thr		
	55	60 65
taagaataaq caagtcttct ttccaggtca ccacttgc 506		

```
<210> 636
<211> 566
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 268..507
```

```
<220>
<221> sig_peptide
<222> 268..363
<223> Von Heijne matrix
score 8.899999961853027
seq LGLFFGFLLVIFA/IE
```

```

<400> 636
acttttagctc gcctaggatt tgagctgggt gtatgtctgc tcccagctca agtcctccg      60
agtgccagag aggaaggcag ggagaagcgg asaccctct ttgggccaaag gccaaggagg      120
actgtgcttg cggggattgc tgtccttgcc attggactat ggctccgatt cgactctcag      180
accaagagca tcttcgagca agaaactaat aataataatt ccagcttcta cacaggagtc      240
tatattctga tcggagccgg cgccctc atg atg ctg gtg ggc ttc ctg ggc tgc      294
                Met Met Leu Val Gly Phe Leu Gly Cys

```

348
-30
-25

tgc ggg gct gtg cag gag tcc cag tgc atg ctg gga ctg ttc ttc ggc	342
Cys Gly Ala Val Gln Glu Ser Gln Cys Met Leu Gly Leu Phe Phe Gly	
-20 -15 -10	
ttc ctc ttg gtg ata ttc gcc att gaa ata gct gcg gcc atc tgg gga	390
Phe Leu Leu Val Ile Phe Ala Ile Glu Ile Ala Ala Ala Ile Trp Gly	
-5 1 5	
tat tcc cac aag gat gag gtg att aag gaa gtc cag gag ttt tac aag	438
Tyr Ser His Lys Asp Glu Val Ile Lys Glu Val Gln Glu Phe Tyr Lys	
10 15 20 25	
gac acc tac aac aag ctg aaa acc aag gat gag ccc cag cgg gaa acg	486
Asp Thr Tyr Asn Lys Leu Lys Thr Lys Asp Glu Pro Gln Arg Glu Thr	
30 35 40	
ctg aaa sca tcc act atg cgt tgaactgctg tggtttggtt gggggcggtg	537
Leu Lys Xaa Ser Thr Met Arg	
45	
aacagtttat ctcagacatc tgcccaaga	566

<210> 637

<211> 568

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 30..527

<220>

<221> sig_peptide

<222> 30..74

<223> Von Heijne matrix

score 8.60000038146973

seq PLLIICLLPAIEG/KN

<400> 637

actggggcac agtaggagga acccagaag atg ctg cct ctc ctg atc atc tgt	53
Met Leu Pro Leu Leu Ile Ile Cys	
-15 -10	
ctc ctg cct gcc att gaa ggg aag aac tgc ctc cgc tgc tgg cca gaa	101
Leu Leu Pro Ala Ile Glu Gly Lys Asn Cys Leu Arg Cys Trp Pro Glu	
-5 1 5	
ctg tct gcc ttg ata gac tat gac ctg cag atc ctc tgg gtg acc cca	149
Leu Ser Ala Leu Ile Asp Tyr Asp Leu Gln Ile Leu Trp Val Thr Pro	
10 15 20 25	
ggg cca ccc aca gaa ctt tct caa aat cgt gac cat ttg gaa gaa gaa	197
Gly Pro Pro Thr Glu Leu Ser Gln Asn Arg Asp His Leu Glu Glu Glu	
30 35 40	
aca gcc aaa ttc ttc act caa gta cac caa gcc att aaa acg tta cga	245
Thr Ala Lys Phe Phe Thr Gln Val His Gln Ala Ile Lys Thr Leu Arg	
45 50 55	
gat gat aaa aca gta ctt ctg gaa gag atc tac acg cac aag aat ctc	293
Asp Asp Lys Thr Val Leu Leu Glu Glu Ile Tyr Thr His Lys Asn Leu	
60 65 70	
ttt act gag agg ctg aat aag ata tct gat ggg ctg aag gag aag gac	341
Phe Thr Glu Arg Leu Asn Lys Ile Ser Asp Gly Leu Lys Glu Lys Asp	
75 80 85	
ata cag tcc aca ctg aag gtc acc agc tgt gct gac tgc agg act cac	389
Ile Gln Ser Thr Leu Lys Val Thr Ser Cys Ala Asp Cys Arg Thr His	
90 95 100 105	
ttc ctc tcc tgc aat gac ccc act ttc tgc cca gcc agg aac cgg cgg	437
Phe Leu Ser Cys Asn Asp Pro Thr Phe Cys Pro Ala Arg Asn Arg Arg	
110 115 120	
acc tcc ctg tgg gct gtg agt ctc agc agt gct cta ctc ctg gcc ata	485

Thr	Ser	Leu	Trp	Ala	Val	Ser	Leu	Ser	Ser	Ala	Leu	Leu	Leu	Ala	Ile	
			125					130					135			
gct	gga	gat	gtb	wct	ttt	act	ggc	aaa	gga	aga	agg	agg	cag			527
Ala	Gly	Asp	Val	Xaa	Phe	Thr	Gly	Lys	Gly	Arg	Arg	Arg	Gln			
		140						145				150				
taaagcaggaacagggcagc cgcattgtct tccagaagtg a																568

```
<220>
<221> CDS
<222> 56..217
```

<400> 638																	
aaaactgaaa acagaacaaa cacaaacaca aacagagaat caacgaagaa gaaaa atg																58	
																Met	
gtg	aag	gtg	atg	tgg	gtt	tcc	gtt	tta	gct	ctg	gcg	gcg	gcg	att	ctc	106	
Val	Lys	Val	Met	Trp	Val	Ser	Val	Leu	Ala	Leu	Ala	Ala	Ala	Ile	Leu		
				-25					-20					-15			-10
ctt	ttg	acg	gtc	ccg	gtg	gca	gaa	ggg	gtg	acg	tgc	tcg	cct	atg	cag	154	
Leu	Leu	Thr	Val	Pro	Val	Ala	Glu	Gly	Val	Thr	Cys	Ser	Pro	Met	Gln		
				-5					1					5			
ctg	gct	tca	tgt	gcg	gcg	gcg	atg	acg	tca	tct	tcg	ccg	cca	tcg	gag	202	
Leu	Ala	Ser	Cys	Ala	Ala	Ala	Met	Thr	Ser	Ser	Ser	Pro	Pro	Ser	Glu		
				10					15					20			
gct	gtt	gca	caa	agc	tgagagagca	gcagccatgc	ctttgtgggt	acatgaggaa								257	
Ala	Val	Ala	Gln	Ser													
																25	
ccctaccctc	cgccaatacg	ttagctcccc	taacgcaagg	aaagtctcca	acagttgcaa											317	
gattcccttc	ccaagctggt	aaggaaatgt	taatcatgat	taattagtga	caagtttcgc											377	
tgattatagt	qgttaatgct	qgtcttatct	tcgtcaqtac	tac												420	

```
<220>  
<221> CDS  
<222> 276..440
```

<400> 639						
tatctcacgc	gcggcagttc	aagaacatca	gacacgctca	acaaacactt	gttgcttgac	60
tgatacatta	tgggcgggcgg	gcaaggcccc	gatcacgtga	taccgcccc	caactccggg	120
cggaagaggc	cagcgacagc	ttgcacgaag	cgagtgggcg	ggcgctcgtg	cccgctaaa	180
aqatgqctgc	tgccgccttg	cagccaccgc	ctgggaggtt	actgtaaggc	ccgcgactcc	240

350

cgccagctcc cgcggactgc tgccgcctcc ttacc atg aag cca gta agt cgt 293
Met Lys Pro Val Ser Arg
-25
cgc acg ctg gac tgg att tat tca gtg ttg ctg ctt gcc atc gtt tta 341
Arg Thr Leu Asp Trp Ile Tyr Ser Val Leu Leu Leu Ala Ile Val Leu
-20 -15 -10 -5
atc tcc tgg ggc tgc atc atc tat gct tgc atg gtg tct gca aga cga 389
Ile Ser Trp Gly Cys Ile Ile Tyr Ala Ser Met Val Ser Ala Arg Arg
1 5 10
cag cta agg aag aaa tac cca gac aaa atc ttt ggg acg aat gaa aat 437
Gln Leu Arg Lys Lys Tyr Pro Asp Lys Ile Phe Gly Thr Asn Glu Asn
15 20 25
ttg taactcttct ggattttaatt atctgaaaat acagttcttt ccctcatgct 490
Leu
tatgtagata taaaaataaa attcataatg caaagt 526

<210> 640

<211> 543

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 152..454

<220>

<221> sig_peptide

<222> 152..217

<223> Von Heijne matrix

score 7.90000009536743

seq ILLVALQARLSHS/RQ

<400> 640

aggaaattag gacacctcat cccaaaagac ctttaaataag gggaagtcca cttgtgcacg 60
gctgctcctt gctatagaag acctgggaca gaggactgct gtctgccctc tctggtcacc 120
ctgcctagct agaggatctg tgaccccgac c atg agg acc ctc gcc atc ctt 172
Met Arg Thr Leu Ala Ile Leu
-20
gct gcc att ctc ctg gtg gcc ctg cag gcc agg ctg agc cac tcc agg 220
Ala Ala Ile Leu Leu Val Ala Leu Gln Ala Arg Leu Ser His Ser Arg
-15 -10 -5 1
caa gag ctg atg agg ttg ctg cag ccc cgg agc aga ttg cag cgg aca 268
Gln Glu Leu Met Arg Leu Leu Gln Pro Arg Ser Arg Leu Gln Arg Thr
5 10 15
tcc cag aag tgg ttg ttt ccc ttg cat ggg acg aaa gct tgg ctc caa 316
Ser Gln Lys Trp Leu Phe Pro Leu His Gly Thr Lys Ala Trp Leu Gln
20 25 30
agc atc cag gct caa gga aaa aca tgg cct gct att gca gaa tac cag 364
Ser Ile Gln Ala Gln Gly Lys Thr Trp Pro Ala Ile Ala Glu Tyr Gln
35 40 45
cgt gca ttg cag gag aac gtc gct atg gaa cct gca tct acc agg gaa 412
Arg Ala Leu Gln Glu Asn Val Ala Met Glu Pro Ala Ser Thr Arg Glu
50 55 60 65
gac tct ggg cat tct gct gct gag ctt gca gaa aaa gaa aaa 454
Asp Ser Gly His Ser Ala Ala Glu Leu Ala Glu Lys Glu Lys
70 75
tgagctcaaa atttgctttg agagctacag ggaattgcta ttactcmtgt accttctgct 514
caatttcctt tctcatctc aaataaatg 543

<210> 641

<211> 336

<212> DNA

<213> Homo sapiens

<220>
 <221> CDS
 <222> 77..247

<220>
 <221> sig_peptide
 <222> 77..139
 <223> Von Heijne matrix
 score 7.80000019073486
 seq SVLIFCLLNLAIS/DR

<220>
 <221> misc_feature
 <222> 301
 <223> n=a, g, c or t

<400> 641
 atctgagcta gggaagaatg tgtattctgt tgttggatga agtagtctac agatgtcagt 60
 tatatccagt tgagta atg gtg ttg ttg agt tca aca atg tct gta ctg att 112
 Met Val Leu Leu Ser Ser Thr Met Ser Val Leu Ile
 -20 -15 -10
 ttc tgc ctg cta aat ctg gcc att tct gat aga agg gta atg aag tct 160
 Phe Cys Leu Leu Asn Leu Ala Ile Ser Asp Arg Arg Val Met Lys Ser
 -5 1 5
 cca act ata ata gtg att cat cta ttt ttt cct cac agt tct acc agt 208
 Pro Thr Ile Ile Val Ile His Leu Phe Phe Pro His Ser Ser Thr Ser
 10 15 20
 att tgc att gag aat ttg gtc act ctg tta tta ggc aca taagaatcct 257
 Ile Cys Ile Glu Asn Leu Val Thr Leu Leu Leu Gly Thr
 25 30 35
 cagctgcatt ttgcattatt tccagagttc ttggttgtaa tcantggata tacaggctgt 317
 ggaagactta ccccatcttc 336

<210> 642
 <211> 507
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 110..496

<220>
 <221> sig_peptide
 <222> 110..187
 <223> Von Heijne matrix
 score 6.59999990463257
 seq LFWLASGWTFAFA/YS

<400> 642
 aagatgcgga ggaggggtgac gcaactagctc tccagttcgc ccggttcctgg cctgaccccc 60
 accaaggccc ataccgcagt aggctcctcg ggctgcccct cggttgaca atg gtc tcc 118
 Met Val Ser
 -25
 agg atg gtc tct acc atg cta tct ggc cta ctg ttt tgg ctg gca tct 166
 Arg Met Val Ser Thr Met Leu Ser Gly Leu Leu Phe Trp Leu Ala Ser
 -20 -15 -10
 gga tgg act cca gca ttt gct tac agc ccc cgg acc cct gac cgg gtc 214
 Gly Trp Thr Pro Ala Phe Ala Tyr Ser Pro Arg Thr Pro Asp Arg Val
 -5 1 5
 tca gaa gca gat atc cag agg ctg ctt cat ggt gtt atg gag caa ttg 262
 Ser Glu Ala Asp Ile Gln Arg Leu Leu His Gly Val Met Glu Gln Leu

352																
10				15				20				25				
ggc	att	gcc	agg	ccc	cga	gtg	gaa	tat	cca	gct	cac	cag	gcc	atg	aat	310
Gly	Ile	Ala	Arg	Pro	Arg	Val	Glu	Tyr	Pro	Ala	His	Gln	Ala	Met	Asn	
				30				35				40				
ctt	gtg	ggc	ccc	cag	agc	att	gaa	ggt	gga	gct	cat	gaa	gga	ctt	cag	358
Leu	Val	Gly	Pro	Gln	Ser	Ile	Glu	Gly	Gly	Ala	His	Glu	Gly	Leu	Gln	
				45				50				55				
cat	ttg	ggt	cct	ttt	ggc	aac	atc	ccc	aac	atc	gtg	gca	gag	ttg	act	406
His	Leu	Gly	Pro	Phe	Gly	Asn	Ile	Pro	Asn	Ile	Val	Ala	Glu	Leu	Thr	
				60				65				70				
gga	gac	aac	att	cct	aag	gac	ttt	agt	gag	gat	cag	ggg	tac	cag	acc	454
Gly	Asp	Asn	Ile	Pro	Lys	Asp	Phe	Ser	Glu	Asp	Gln	Gly	Tyr	Gln	Thr	
				75				80				85				
ctc	caa	atc	cct	gtc	ctg	ttg	gaa	aaa	cag	atg	atg	gat	gtc			496
Leu	Gln	Ile	Pro	Val	Leu	Leu	Glu	Lys	Gln	Met	Met	Asp	Val			
				90				95				100				
tagaaaacac c																507

```
<210> 643
<211> 466
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 14..310
```

```
<220>
<221> sig_peptide
<222> 14..136
<223> Von Heijne matrix
      score 5.400000009536743
      seq WLCPGLRLPGSLA/GR
```

<400>	643	
gtaaacagac	aac atg gcg gcc gcg gtg kss gcg gca cct ggg gcc ttg	49
	Met Ala Ala Ala Val Xaa Ala Ala Pro Gly Ala Leu	
	-40 -35 -30	
gga tcc ctg cat gct ggc ggc gcc cgc ctg gtg gcc gct tgc agt gcg		97
Gly Ser Leu His Ala Gly Gly Ala Arg Leu Val Ala Ala Cys Ser Ala		
	-25 -20 -15	
tgg ctc tgc ccg ggg ttg agg ctg ccc ggc tcg ttg gca ggc cgg cga		145
Trp Leu Cys Pro Gly Leu Arg Leu Pro Gly Ser Leu Ala Gly Arg Arg		
	-10 -5 1	
gcc ggc ccg gcg atc tgg gcc cag ggc tgg gta cct gcg gcc ggg ggt		193
Ala Gly Pro Ala Ile Trp Ala Gln Gly Trp Val Pro Ala Ala Gly Gly		
	5 10 15	
ccc gcc ccg aaa agg ggc tac agc tct gag atg aag acg gag gac gag		241
Pro Ala Pro Lys Arg Gly Tyr Ser Ser Glu Met Lys Thr Glu Asp Glu		
	20 25 30 35	
ctg cgg gtg cgg cac ctg gag gag gag aac cga gac aga gac aag aac		289
Leu Arg Val Arg His Leu Glu Glu Glu Asn Arg Asp Arg Asp Lys Asn		
	40 45 50	
atg gtc aaa ggc atg gag atg tgaacatgtg tggtgtattt gggacgtaac		340
Met Val Lys Gly Met Glu Met		
	55	
ataaggaatt gtggtgcttg gaataaacag agcttatggc aaaaattcac tcagtaaaaa		400
tcttataaaa atgctatcaa aagctgtgga tgctttgaaa tctgataaga aagtacggac		460
cataat		466

<210>	644
<211>	502
<212>	DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 19..342

<220>

<221> sig_peptide

<222> 19..75

<223> Von Heijne matrix

score 4.19999980926514

seq AVAASAASGQAEK/KK

<400> 644

```

gcgcgctgct cttctaag atg gct gcc gct acc ggt gcg gtg gca gcc tcg      51
                        Met Ala Ala Ala Thr Gly Ala Val Ala Ala Ser
                        -15                                -10
gcc gcc tcg ggt cag gcg gaa ggt aaa aag atc acc gat ctg cgg gtc      99
Ala Ala Ser Gly Gln Ala Glu Gly Lys Lys Ile Thr Asp Leu Arg Val
                        -5                                1          5
atc gat ctg aag tcc gag ctg aag cgg cgg aac tta gac atc acc gga      147
Ile Asp Leu Lys Ser Glu Leu Lys Arg Arg Asn Leu Asp Ile Thr Gly
                        10                                15          20
gtc aag acc gtg ctc atc tcc cga ctc aag cag gct att gaa gag gaa      195
Val Lys Thr Val Leu Ile Ser Arg Leu Lys Gln Ala Ile Glu Glu Glu
25                                30                                35          40
gga ggc gat cca gat aat att gaa tta act gtt tca act gat act cca      243
Gly Gly Asp Pro Asp Asn Ile Glu Leu Thr Val Ser Thr Asp Thr Pro
                        45                                50          55
aac aag aaa cca act aaa ggc aaa ggt aaa aaa cat gaa gca gat gag      291
Asn Lys Lys Pro Thr Lys Gly Lys Gly Lys Lys His Glu Ala Asp Glu
60                                65                                70
ttg agt gga gat gct tct gtg gaa gat gat gct ttt atc aag gta aag      339
Leu Ser Gly Asp Ala Ser Val Glu Asp Asp Ala Phe Ile Lys Val Lys
75                                80                                85
tgt taattactca ggттаagatg cttattttaa gcctagtatt ttgcatcatg      392
Cys
aattttgtgc tttttgttga acaatgggtca ctgaagtatg taggttaaag akdaagacct      452
atgtaatat gacaacgttt aatattgtca atctcattaa ctctcataat      502

```

<210> 645

<211> 493

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 264..443

<220>

<221> sig_peptide

<222> 264..386

<223> Von Heijne matrix

score 4.19999980926514

seq VFWSFYLSRVNC/TT

<400> 645

```

ttcgtgaaat ttgttggttt tgaggactgt aaaagtgatt tcatactctg aatataaaac      60
tggaataatag ggtaatgttt taaaatttat tatgctatta ttcagaatgc caaagtatta      120
ttttttttcc caaaatcagt ctggacatct actacttttt agactttttg acgttgaact      180
ttctgtataa aaattggctg ggttttgagc ttttggttaag aaataaaaagc cgattaagca      240
ctggccgccc cgcgtggtac cca atg ccc gag tca ctg tgg cag cat tcg cac      293
                        Met Pro Glu Ser Leu Trp Gln His Ser His

```

354

```

                -40                -35
tgg tgt ggg gag tcc ttt caa ctc aag gag gct ggg ttt ctg ggc acc      341
Trp Cys Gly Glu Ser Phe Gln Leu Lys Glu Ala Gly Phe Leu Gly Thr
      -30                -25                -20
ctt gaa gtt ttc tgg atg tct ttt tat ctt tct cgt gtg aac tgc act      389
Leu Glu Val Phe Trp Met Ser Phe Tyr Leu Ser Arg Val Asn Cys Thr
-15                -10                -5                1
aca aaa gag acc agc cca ctt ccc aag cca gcc aga cac ctg ggt ctt      437
Thr Lys Glu Thr Ser Pro Leu Pro Lys Pro Ala Arg His Leu Gly Leu
      5                10                15
gag cca taaactggcg tagttaagct ttgcagcttc cagtgtatatt tattttattct      493
Glu Pro

```

<210> 646
 <211> 540
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 175..354

<220>
 <221> sig_peptide
 <222> 175..309
 <223> Von Heijne matrix
 score 4
 seq ENFLSLLSKSCSA/DP

```

<400> 646
attttaaaat ggaaattggg gcgagggggtt ggcggtctggg cgaagggagc attcaaaagc      60
ggagaatgtc actttaccgc agaaattcac tacgatgggc cattgggttc tgctcggctc      120
gcaggcgcac tgcaagagtg gaggtgtggc tagtggctgt gatgagataa atcc atg      177
                                         Met
                                         -45
cat agc ctt ttc att gcg agc ttg aaa gtt ctt ttt tat tac agt ttt      225
His Ser Leu Phe Ile Ala Ser Leu Lys Val Leu Phe Tyr Tyr Ser Phe
      -40                -35                -30
agc ttt agg ttt aat tgg ttc gac tgc ctt ctc cac aat ttg ggc gag      273
Ser Phe Arg Phe Asn Trp Phe Asp Cys Leu Leu His Asn Leu Gly Glu
      -25                -20                -15
aat ttc ctt agc ctt ctc agc aaa agt tgt tct gcg gac ccg tct ggg      321
Asn Phe Leu Ser Leu Leu Ser Lys Ser Cys Ser Ala Asp Pro Ser Gly
      -10                -5                1
tca act ttc atg agg gac att gag aca aac aaa tgaaatatgg gttaaagtac      374
Ser Thr Phe Met Arg Asp Ile Glu Thr Asn Lys
5                10                15
tctgagcagc tacaaaaaga agaccagtct atcctgctgg agacagtggc cacgtgaaga      434
aagagctctt gcagtatgaa gaccacatgg aaagagaggc cacatggaac caacagtcag      494
catcttggtt tcggacacgt gaagaattca tctcagactg tgtatc      540

```

<210> 647
 <211> 505
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 100..498

<220>
 <221> sig_peptide
 <222> 100..138

<223> Von Heijne matrix
 score 3.79999995231628
 seq MAHCVTLVQLSIS/CD

<400> 647

```

attttgtgaa gcggcgaagg aggtggtggc tgcgttgggc tccgggaagc cgttcgggct      60
ggggctgtcg gccgcggggc ggaggcactc gcgcggggg atg gcc cac tgc gtg      114
                                   Met Ala His Cys Val
                                   -10
acc ttg gtt cag ctg tcc att tcc tgt gac cat ctc att gac aag gac      162
Thr Leu Val Gln Leu Ser Ile Ser Cys Asp His Leu Ile Asp Lys Asp
                                   -5                                   5
atc ggc tcc aag tct gac cca ctc tgc gtc ctt tta cag gat gtg gga      210
Ile Gly Ser Lys Ser Asp Pro Leu Cys Val Leu Leu Gln Asp Val Gly
                                   10                                   15                                   20
ggg ggc agc tgg gct gag ctt ggc cgg act gaa cgg gtg cgg aac tgc      258
Gly Gly Ser Trp Ala Glu Leu Gly Arg Thr Glu Arg Val Arg Asn Cys
25                                   30                                   35                                   40
tca agc cct gag ttc tcc aag act cta cag ctt gag tac cgc ttt gag      306
Ser Ser Pro Glu Phe Ser Lys Thr Leu Gln Leu Glu Tyr Arg Phe Glu
                                   45                                   50                                   55
aca gtc cag aag cta cgc ttt gga atc tat gac ata gac aac aag acg      354
Thr Val Gln Lys Leu Arg Phe Gly Ile Tyr Asp Ile Asp Asn Lys Thr
                                   60                                   65                                   70
cca gag ctg agg gat gat gac ttc cta ggg ggt gct gag tgt tcc cta      402
Pro Glu Leu Arg Asp Asp Asp Phe Leu Gly Gly Ala Glu Cys Ser Leu
                                   75                                   80                                   85
gga cag att gtg tcc agc cag gta ctg act ctc ccc ttg atg ctg aag      450
Gly Gln Ile Val Ser Ser Gln Val Leu Thr Leu Pro Leu Met Leu Lys
90                                   95                                   100
cct gga aaa cct gct ggg cgg ggg aca tca cgg tct cag ctc agg aat      498
Pro Gly Lys Pro Ala Gly Arg Gly Thr Ser Arg Ser Gln Leu Arg Asn
105                                   110                                   115                                   120
taaagga
                                         505

```

<210> 648

<211> 472

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 78..281

<220>

<221> sig_peptide

<222> 78..218

<223> Von Heijne matrix
 score 3.70000004768372
 seq PLLSLHSRGGSSS/ES

<400> 648

```

attttgcgaa cggcgagcag cggcgcggcg cggagagacs agcggagggtt ttctgtgttt      60
cggaccccag cggccgg atg gtg aaa tcc tcc ctg cag cgg atc ctc aat      110
                                   Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn
                                   -45                                   -40
agc cac tgc ttc gcc aga gag aag gaa ggg gat aaa ccc agc gcc acc      158
Ser His Cys Phe Ala Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr
                                   -35                                   -30                                   -25
atc cac gcc agc cgc acc atg ccg ctc cta agc ctg cac agc cgc ggc      206
Ile His Ala Ser Arg Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly
-20                                   -15                                   -10                                   -5
ggc agc agc agt gag agt tcc agg gtc tcc ctc cac tgc tgt agt aac      254

```

```

Gly Ser Ser Ser Ser Glu Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn
      1      5      10
ccg ggt ccg ggg cct cgg tgg tgc tcc tgatgccct caccacccc 301
Pro Gly Pro Gly Pro Arg Trp Cys Ser
      15      20
tgaagatccc aggtgggcga gggaatagtc agaggggatca caatctttca gctttttggg 361
ctttgagatt gtgagaccgg ggcacccct tatccccaag agaccgacg cttgcttcat 421
ggcctacacg ttcgagagag agtcttcggg agaggaggag gaggtagggcc g 472

<210> 649
<211> 458
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 173..373

<220>
<221> sig_peptide
<222> 173..265
<223> Von Heijne matrix
      score 3.70000004768372
      seq ASWALGRWGSCRA/LD

<400> 649
agaatctgac attccaggac tccactagtc ctgttttcct ctgtgggaac tagagccaag 60
gcgagagacc cgtgccagcc ccgaggctcc cggggcccat gggcaggcgg ggctggtaca 120
gggccccggg gctctctcag cccatctgtc acccctcccc cccaacaccc ag atg tcc 178
                                         Met Ser
                                         -30
cgt ctt ggc agt gct gtc ccg gac tgc cac att aat gct cca gaa ccc 226
Arg Leu Gly Ser Ala Val Pro Asp Cys His Ile Asn Ala Pro Glu Pro
      -25      -20      -15
gcc agc tgg gcc ctg ggg cgg tgg ggc tcc tgc aga gcc ctg gac aaa 274
Ala Ser Trp Ala Leu Gly Arg Trp Gly Ser Cys Arg Ala Leu Asp Lys
      -10      -5      1
ggc gag gga ggg agg gaa ggg gat cct aag cac ccc tcc ctc cct ggc 322
Gly Glu Gly Gly Arg Glu Gly Asp Pro Lys His Pro Ser Leu Pro Gly
      5      10      15
cct agg agg cag aca tgc ccg atg aga tca aca ttg atg aat tgt tgg 370
Pro Arg Arg Gln Thr Cys Pro Met Arg Ser Thr Leu Met Asn Cys Trp
      20      25      30      35
agt tagagagtga agaggagaga agccggaaaa tccagggact cctgaagtca 423
Ser
tgtgggaaac ctgtcgagga cttcatccag ggagc 458

<210> 650
<211> 482
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 8..250

<220>
<221> sig_peptide
<222> 8..64
<223> Von Heijne matrix
      score 3.5
      seq RLLRASVARHVSA/IP

```

357

<400> 650
 actgaag atg gcg gct gct gta gga cgg ttg ctc cga gcg tcg gtt gcc 49
 Met Ala Ala Ala Val Gly Arg Leu Leu Arg Ala Ser Val Ala
 -15 -10
 cga cat gtg agt gcc att cct tgg ggc att tct gcc act gca gcc ctc 97
 Arg His Val Ser Ala Ile Pro Trp Gly Ile Ser Ala Thr Ala Ala Leu
 -5 1 5 10
 agg cct gct gca tgt gga aga acg agc ttg aca aat tta ttg tgt tct 145
 Arg Pro Ala Ala Cys Gly Arg Thr Ser Leu Thr Asn Leu Leu Cys Ser
 15 20 25
 ggt tcc agt caa gca aaa att att tct tct atc ctt tgg att tca cct 193
 Gly Ser Ser Gln Ala Lys Ile Ile Ser Ser Ile Leu Trp Ile Ser Pro
 30 35 40
 ttg tgt gtc cta cag aaa ttg ttg ctt tta gtg aca aag cta acg aat 241
 Leu Cys Val Leu Gln Lys Leu Leu Leu Leu Val Thr Lys Leu Thr Asn
 45 50 55
 ttc acg atg tgaactgtga agttgtcgca gtctcagtgg attcccactt 290
 Phe Thr Met
 60
 tagccatctt gcctggataa atacaccaag aaagaatggt ggtttgggcc acatgaacat 350
 cgcaactcttg tcagacttaa ctaagcagat ttcccgagac tacggtgtgc tgttagaagg 410
 ttctggtctt gcaactaagag gtctcttcat aattgaccct aatggagtca tcaagcattt 470
 gagcgtcaac ga 482

<210> 651
 <211> 291
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 130..288

<220>
 <221> sig_peptide
 <222> 130..234
 <223> Von Heijne matrix
 score 3.5
 seq GLVEVVAVGVVRS/DQ

<400> 651
 agtctctgtc ggagttctgtt tcccagtgca cccaacttgt gatagaaggt atatgttctg 60
 ctgaaaaaga gaaaaaggaa aaaaaagaat aaaagcagag agcagagaag gacaggtct 120
 ggctgccgc atg gag ggg agg agc agg agc agg aaa gcc tgt ctg gag gtt 171
 Met Glu Gly Arg Ser Arg Ser Arg Lys Ala Cys Leu Glu Val
 -35 -30 -25
 gct aca gta aac cag gtc ttc gtg ggg ttg gtc gag gtg gtg gca gtg 219
 Ala Thr Val Asn Gln Val Phe Val Gly Leu Val Glu Val Val Ala Val
 -20 -15 -10
 ggg gtg gtg aga agt gat cag att aag aat atg tat tta aaa tac agc 267
 Gly Val Val Arg Ser Asp Gln Ile Lys Asn Met Tyr Leu Lys Tyr Ser
 -5 1 5 10
 cag cat gat ctg ctg atg gac tag 291
 Gln His Asp Leu Leu Met Asp
 15

<210> 652
 <211> 766
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS

<222> 166..333

<400> 652

```

aaaaggagta gctattagcc aattcggcag ggcccgcttt ttagaagctt gatttcyttt      60
gaagatgaaa gactagcgga agctctgcct ctttccccag tggcgaggga actcggggcg      120
attggctggg aactgtatcc acccaaatgt caccgatttc ttcct atg cag gaa atg      177
                               Met Gln Glu Met
                               1
agc aga ccc atc aat aag aaa ttt ctc agc ctg gcc gaa aat ggt tgg      225
Ser Arg Pro Ile Asn Lys Lys Phe Leu Ser Leu Ala Glu Asn Gly Trp
5                               10                               15                               20
ccc cac gaa gcc acg aca act gga ggc aaa gag ggt tgc tca acg ccc      273
Pro His Glu Ala Thr Thr Thr Gly Gly Lys Glu Gly Cys Ser Thr Pro
                               25                               30                               35
cgc ctc att gga aaa cca aat cag atc tgg gac cta tat agc gtg gcg      321
Arg Leu Ile Gly Lys Pro Asn Gln Ile Trp Asp Leu Tyr Ser Val Ala
                               40                               45                               50
gag gcg ggg cga tgattgtcgc gctcgcaccc actgcagctg cgcacagtcg      373
Glu Ala Gly Arg
55
catttctttc cccgcccctg agaccctgca gcaccatctg tcatggcggc tgggctgttt      433
ggtttgagcg ctccgcgtct tttggcggca gcggcgacgc gagggctccc ggccgcccgc      493
gtccgctggg aatctagctt ctccaggact gtggtcgccc cgtcgcgtgt gccggaagc      553
ggccccaga accgaccaca ccgtgcaaga ggaccagaa cccgaggacg aaaacttgta      613
tgagaagaac ccagactccc atggttatga caaggacccc gttttggacg tctggaacat      673
gcgacttgtc ttcttctttg gcgtctccat catcctggtc cttggcagca cctttgtggc      733
ctatctgcct gactacagga tgaaagagtg ggt                                766

```

<210> 653

<211> 488

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 199..357

<400> 653

```

cagtaggata catagctttt cttccagtga aacaaagttc atatcatcca ttgtttttca      60
agcacgtgac accagcctca aagtaaata catgaccagt ggttgaacag tctaattttc      120
aaatttaata tagagcatat aacttctgat ttgatagtat ttattttaaa aaattatgtt      180
ttcatcattc atttgaaa atg aaa aag ccc caa agt gag aac ttt ggg gga      231
                               Met Lys Lys Pro Gln Ser Glu Asn Phe Gly Gly
                               1                               5                               10
ggg cct aga aca tgg ata gat ctc tta gtg gtc ttt cca aaa gta cat      279
Gly Pro Arg Thr Trp Ile Asp Leu Leu Val Val Phe Pro Lys Val His
                               15                               20                               25
gta ctt gaa ata ttt tca tta tca tac tat tct ttg aaa aaa aag atg      327
Val Leu Glu Ile Phe Ser Leu Ser Tyr Tyr Ser Leu Lys Lys Lys Met
                               30                               35                               40
ctt act gta tac ttg ttt tca agc atc ctc taaaatcaaa ggttttgatc      377
Leu Thr Val Tyr Leu Phe Ser Ser Ile Leu
                               45                               50
acaatatgca gatttctctt gatagatact taaataggct atttctctcc tcttcttggg      437
caatgccttg ttttctcctc tgaatatattg catttgaaag gattgcttcc t                                488

```

<210> 654

<211> 489

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 253..426

<400> 654

```

aacctcacga ggccttacgt gcctagagcg gttcctctcc gtggcgccct accgccttcg      60
gcgcgaaggt ggctgggtgcg taccgtcccg aaccctagga gccgagatcc caccgcagg      120
accccaaaac ccaatgatcc tgcagcagcc cttgcagcga ggcccccagg gaggggcccag      180
cgctcccgcg gggcgccctt ggggggtgact tggggcctgg acgccagctc csctctscga      240
ggagctgtgc cc atg agc acc aag cgg cgc ctg gag gag gag cag gag cct      291
              Met Ser Thr Lys Arg Arg Leu Glu Glu Glu Gln Glu Pro
                1          5          10
ctg cgc aag cag ttt ctg tct gag gag aac atg gcc acc cac ttc tct      339
Leu Arg Lys Gln Phe Leu Ser Glu Glu Asn Met Ala Thr His Phe Ser
   15          20          25
caa ctc agc ctg cac aat gac caa ccc cta ctg cag ccc cac cat gac      387
Gln Leu Ser Leu His Asn Asp Gln Pro Leu Leu Gln Pro His His Asp
   30          35          40          45
ctt ctc ccc agc cct gcc ccc act cag gag ccc ttg ctc tgagctgctt      436
Leu Leu Pro Ser Pro Ala Pro Thr Gln Glu Pro Leu Leu
              50          55
ctctggcgct atcctggcag cctcatcctg aggcctccgt ctgctgaggc tgg      489

```

<210> 655

<211> 616

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 205..387

<400> 655

```

acttcggggg gagcgggcgcg gcggcgggga ggatctctca ccccatcact cagggtggcg      60
caatcacgac tcattggctc actgcagcct agacctcca gctggagcaa ttctcctgcc      120
tcagccttct gagtagctgg gactacagtt ggttctaaag agtggtgagt cagaagagac      180
gtcaggcagc aagcgacttg ggcc atg gcc tct gac cta gac ttc tca cct      231
              Met Ala Ser Asp Leu Asp Phe Ser Pro
                1          5
ccg gag tgc ccg agc cca ctt tcc tgg aga acc tgc tac ggt acg gac      279
Pro Glu Cys Pro Ser Pro Leu Ser Trp Arg Thr Cys Tyr Gly Thr Asp
   10          15          20          25
tct tcc tgg gag cca tct tcc agc tca tct gtg tgc tgg cca tca tcg      327
Ser Ser Trp Glu Pro Ser Ser Ser Ser Val Cys Trp Pro Ser Ser
              30          35          40
tac cca ttc cca agt ccc acg agg cgg agg ctg aac cgt ctg agc cca      375
Tyr Pro Phe Pro Ser Pro Thr Arg Arg Leu Asn Arg Leu Ser Pro
              45          50          55
gaa gtg ctg agg tgacgaggaa gcccaaggct gctgttcctt ctgtgaacaa      427
Glu Val Leu Arg
   60
gaggcccaag aaagagacta agaagaagcg gtagaagagg aggcctgagg actgggcgkg      487
caggagagg gtcttkgggg acagcctcct gggaatctac attgtgtttc cccgcatty      547
ccaggctcag gttctgarga rgctgtgacg ccctatgacc gcagagatct agacagtcgt      607
aacagtccc                                     616

```

<210> 656

<211> 508

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 37..219

360

<400> 656

```

ctttcccctg ccggtgtcct gctcgccgtc cccgcc atg ctg tct cta gac ttt      54
                                   Met Leu Ser Leu Asp Phe
                                   1      5
ttg gac gat gtg cgg cgg atg aac aag cgg cag ctc tat tat caa gtc      102
Leu Asp Asp Val Arg Arg Met Asn Lys Arg Gln Leu Tyr Tyr Gln Val
                                   10      15      20
cta aat ttt gga atg att gtc tca tcg gca cta atg atc tgg aag ggg      150
Leu Asn Phe Gly Met Ile Val Ser Ser Ala Leu Met Ile Trp Lys Gly
                                   25      30      35
tta atg gta ata act gga agt gaa agt ccg att gta gtg gtc tca ggc      198
Leu Met Val Ile Thr Gly Ser Glu Ser Pro Ile Val Val Val Ser Gly
                                   40      45      50
aaa atg ggc ata tca agt ttt tgaccaaagg agataataat gcggttgatg      249
Lys Met Gly Ile Ser Ser Phe
55      60
accgaggcct ctataaaca ggacaacatt ggctagagaa aaaagatggt gtggggagag      309
ccaggggatt tgttccttat attggaattg tgacgatcct catgaatgac taccctaaat      369
ttaagtatgc agttctcttt ttgctgggtt tattcgtgct ggttcacgt gagtaagaag      429
cctgccttgc tgttcctggg aagatgccat agttttcgtt actggatggt tggagtagat      489
actggtctgt gattggtgg      508

```

<210> 657

<211> 500

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 114..452

<400> 657

```

ttttcccctg ccggtgtcct gctcgccgtc cccgccatgc tgtctctaga ctttttggac      60
gatgtgcggc ggatgaacaa gcggcastct attatcaagt cctaaatttt gga atg      116
                                   Met
                                   1
att gtc tca tcg gca cta atg atc tgg aag ggg tta atg gta ata act      164
Ile Val Ser Ser Ala Leu Met Ile Trp Lys Gly Leu Met Val Ile Thr
                                   5      10      15
gga agt gaa agt ccg att gta gtg gtg ctc agt ggc agc atg gaa cct      212
Gly Ser Glu Ser Pro Ile Val Val Val Leu Ser Gly Ser Met Glu Pro
                                   20      25      30
gca ttt cat aga gga gat ctt ctc ttt cta aca aat cga gtt gaa gat      260
Ala Phe His Arg Gly Asp Leu Leu Phe Leu Thr Asn Arg Val Glu Asp
                                   35      40      45
ccc ata cga gtg gga gaa att gtt gtt ttt agg ata gaa gga aga gag      308
Pro Ile Arg Val Gly Glu Ile Val Val Phe Arg Ile Glu Gly Arg Glu
50      55      60      65
att cct ata gtt cac cga gtc ttg aag att cat gaa aaa ttt gtt cct      356
Ile Pro Ile Val His Arg Val Leu Lys Ile His Glu Lys Phe Val Pro
                                   70      75      80
tat att gga att gtg acg atc ctc atg aat gac tat cct aaa ttt aag      404
Tyr Ile Gly Ile Val Thr Ile Leu Met Asn Asp Tyr Pro Lys Phe Lys
                                   85      90      95
tat gca gtt ctc ttt ttg ctg ggt tta ttc gtg ctg gtt cat cgt gag      452
Tyr Ala Val Leu Phe Leu Leu Gly Leu Phe Val Leu Val His Arg Glu
100      105      110
taagaagcct gccttgctgt tcctggggaa gatgccatag ttttcggt      500

```

<210> 658

<211> 411

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 90..305

<220>

<221> misc_feature

<222> 138,315,326,343,360,377

<223> n=a, g, c or t

<400> 658

```

ttgggcagaa aaattcaagc aggagattgt atttcctttg gagttgtacg atttccttta      60
ttatttgaac tgcagtaaag aaagctggg atg ggc tcc tct agg gat act tcc      113
                               Met Gly Ser Ser Arg Asp Thr Ser
                               1           5

aga tcc ctg ggc ggt tgt agc cct ngc tcc tct tta aat gga ttt ggt      161
Arg Ser Leu Gly Gly Cys Ser Pro Xaa Ser Ser Leu Asn Gly Phe Gly
   10           15           20

ttc aaa gac gat cat ctc cgt ctt ctc gga tgt cat agt gcc act gat      209
Phe Lys Asp Asp His Leu Arg Leu Leu Gly Cys His Ser Ala Thr Asp
   25           30           35           40

cat ctc cag ctc ctk gcc acc ctg ggc ttt ctc cac ttt tgc ctc tat      257
His Leu Gln Leu Leu Ala Thr Leu Gly Phe Leu His Phe Cys Leu Tyr
           45           50           55

gtt ttg ctt ctc cam cgt ctt agc cac gat atc tac ctc tct gtc atg      305
Val Leu Leu Leu Xaa Arg Leu Ser His Asp Ile Tyr Leu Ser Val Met
           60           65           70

tgatgtgacn cttgtttttt naaccaagag ttgccctnaa gstcccttaa aaaanagctg      365
atcttactgg snttcttttg taaarctcct gtgatagatg cagatt      411

```

<210> 659

<211> 529

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 235..444

<400> 659

```

atttccagcg gttgctgggt ctgacggggt gtagtctgcc aggacaatga gttatgacta      60
ccatcagaac tggggccgtg atgggggtcc ccgcagctcc ggtgggggct atggaggggg      120
gccagcaggg ggtcatggag gtaaccgagg ctccggagga grcggcggcg gcggaggggg      180
tggtcgaggg ggcaggggcc ggcattcccgg gcacctgaaa gccgcgaaat cggc atg      237
                               Met
                               1

tgg tac gcg aaa aaa cag ggg cag aag aac aag gaa gcg gag agg caa      285
Trp Tyr Ala Lys Lys Gln Gly Gln Lys Asn Lys Glu Ala Glu Arg Gln
           5           10           15

gag aga gct gta gta cac atg gat gaa cga cga gaa gaa caa att gta      333
Glu Arg Ala Val Val His Met Asp Glu Arg Arg Glu Glu Gln Ile Val
           20           25           30

cag tta ctg aat tct gtt caa gcg aag aat gat aaa gag tca gaa gca      381
Gln Leu Leu Asn Ser Val Gln Ala Lys Asn Asp Lys Glu Ser Glu Ala
           35           40           45

cag ata tcc acc cca gac ttg gct gtc cgc tct gcg cct tcg ggc cag      429
Gln Ile Ser Thr Pro Asp Leu Ala Val Arg Ser Ala Pro Ser Gly Gln
           50           55           60           65

ccc ttt cgg gct tgc tgagcctgga ggccgagag aatgcacttc cgggttttgc      484
Pro Phe Arg Ala Cys
           70

tgaggctctg aggagctacc aggaggctgc ggctgcaggg acctt      529

```

```
<210> 660
<211> 478
<212> DNA
<213> Homo sapiens
```

```
<220>  
<221> CDS  
<222> 128..286
```

<400>	660	
gttcttttggc cctgtgacac gtagcaacgg ggctggttca gggctctgaaa cagagtttgg		60
gggttggttg ggattagtga agctactgcc tttgccgcc ggcagctca gagtttgatt		120
atttgca atg tca ggc ttt gaa aac tta aac acg gat ttc tac cag aca		169
Met Ser Gly Phe Glu Asn Leu Asn Thr Asp Phe Tyr Gln Thr		
1 5 10		
agt tac agc atc gat gat cag tca cag cag tcc tat gat tat gga gga		217
Ser Tyr Ser Ile Asp Asp Gln Ser Gln Gln Ser Tyr Asp Tyr Gly Gly		
15 20 25 30		
agt gga gga ccc tat aac agt atg ctg gct atg act att cgc agc aag		265
Ser Gly Gly Pro Tyr Asn Ser Met Leu Ala Met Thr Ile Arg Ser Lys		
35 40 45		
gca gat ttg tcc ctc cag aca tgatgcagcc acaacagcca tacaccgggc		316
Ala Asp Leu Ser Leu Gln Thr		
50		
agatttacca gccactcag gcatatactc cagcttcacc tcagcctttc tatggaaaca		376
actttgagga tgagccacct ttattagaag agttagggtat caattttgac cacatctggc		436
aaaaaacact aacagtatta catccgttaa aagtagcaga tg		478

```
<210> 661
<211> 456
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> 110..277
```

<400>	661		
tacggttttta	ttattatatc	gatttatttg tgaggagggga atctgttcac gtagatttta	60
aaagtgc	aag agtgtccagc	gtcatctgcc tctcctgtcc cctgaccac atg att ttc	118
		Met Ile Phe	
		1	
tgt tcc cat	aaa cta gca ctg ttg agt ata ttt cca gag cat act cca		166
Cys Ser His	Lys Leu Ala Leu Leu Ser Ile Phe Pro Glu His Thr Pro		
5	10	15	
ggc gtg tgt	aaa tac agc ctt cca ctc ctg cac aca cac cgt gtt ttt		214
Gly Val Cys	Lys Tyr Ser Leu Pro Leu Leu His Thr His Arg Val Phe		
20	25	30	35
aca tca gta	aca ttc tgc gca tat tgt tca gca cca tgc ctt ttc cac		262
Thr Ser Val	Thr Phe Cys Ala Tyr Cys Ser Ala Pro Cys Leu Phe His		
	40	45	50
tta aca gta	ata att taaaggtcac tccatttcag tacataaaga gcgtcctcag		317
Leu Thr Val	Ile Ile		
	55		
gccgggcgcg	gtggctcgca cctgtaatcc tagcactttg ggaggctgag gcaggtggat		377
cacctgaggt	tggaagttga agaccagcct ggccaaaagg gcgaaaccct gtctctactg		437
aaaataaaaa	aattaaqctc		456

```
<210> 662
<211> 506
<212> DNA
<213> Homo sapiens
```

<220>

<221> CDS

<222> 71..379

<400> 662

```

acagctgccc gggactccag tgatcgccgc ggctcgctcg cgccccggaa actgcccctt      60
ctcgggggtc atg atg ggc agc aag atg gcg tct gct agt agg gtc gtt      109
      Met Met Gly Ser Lys Met Ala Ser Ala Ser Arg Val Val
      1      5      10
cag gta gtc aaa cca cac act cca tta ata agg ttt cct gac aga aga      157
Gln Val Val Lys Pro His Thr Pro Leu Ile Arg Phe Pro Asp Arg Arg
      15      20      25
gac aat cct aaa ccc aat gta tca gaa gct ttg aga tca gca ggg cta      205
Asp Asn Pro Lys Pro Asn Val Ser Glu Ala Leu Arg Ser Ala Gly Leu
      30      35      40      45
cca tct cac tct tct gta att tca caa cat tct aaa gga agt aaa tca      253
Pro Ser His Ser Ser Val Ile Ser Gln His Ser Lys Gly Ser Lys Ser
      50      55      60
cca gat ttg ctg atg tat cag ggt cca cca gac act gca gaa ata ata      301
Pro Asp Leu Leu Met Tyr Gln Gly Pro Pro Asp Thr Ala Glu Ile Ile
      65      70      75
aaa aca tta cct cag aaa tac aga agg aaa ctt gtg tct caa gaa gaa      349
Lys Thr Leu Pro Gln Lys Tyr Arg Arg Lys Leu Val Ser Gln Glu Glu
      80      85      90
atg gaa ttt atc caa cgt gga ggt cct gaa taaccatggt ggctgctggt      399
Met Glu Phe Ile Gln Arg Gly Gly Pro Glu
      95      100
tgtcatcaga caatagaatt gtctttacaa taaaggactt ccaaaatgac agatgagaaa      459
ctgtatatta aacaccttta ataaatatta tgaaaaaaam haawaaa      506

```

<210> 663

<211> 595

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 207..578

<220>

<221> misc_feature

<222> 19

<223> n=a, g, c or t

<400> 663

```

atctcgctgc cccaacggnc ccccgggctcg gtttccgcgg tggccatgac tgcggccgtg      60
ttcttcggct gcgcttcatt gccttcgggc ctgcgctcgc cctttatgtc ttcaccatcg      120
ccaycgagcc gttgcgtatc atcttcctca tcgccggagc tttcttctgg ttggtgtctc      180
tactgatttc gtcccttggt tggttc atg gca aga gtc att att gac aac aaa      233
      Met Ala Arg Val Ile Ile Asp Asn Lys
      1      5
gat gga cca aca cag aaa tat ctg ctg atc ttt gga gcg ttt gtc tct      281
Asp Gly Pro Thr Gln Lys Tyr Leu Leu Ile Phe Gly Ala Phe Val Ser
      10      15      20      25
gtc tat atc caa gaa atg ttc cga ttt gca tat tat aaa ctc tta aaa      329
Val Tyr Ile Gln Glu Met Phe Arg Phe Ala Tyr Tyr Lys Leu Leu Lys
      30      35      40
aaa gcc agt gaa ggt ttg aag agt ata aac cca ggt gag aca gca ccc      377
Lys Ala Ser Glu Gly Leu Lys Ser Ile Asn Pro Gly Glu Thr Ala Pro
      45      50      55
tct atg cga ctg ctg gcc tat gtt tct ggc ttg ggc ttt gga atc atg      425
Ser Met Arg Leu Leu Ala Tyr Val Ser Gly Leu Gly Phe Gly Ile Met
      60      65      70

```

364

```

agt gga gta ttt tcc ttt gtg aat acc cta tct gac tcc ttg ggg cca      473
Ser Gly Val Phe Ser Phe Val Asn Thr Leu Ser Asp Ser Leu Gly Pro
  75                      80                      85
ggc aca gtg ggc att cat gga gat tct cct caa ttc ttc ctt tta ttc      521
Gly Thr Val Gly Ile His Gly Asp Ser Pro Gln Phe Phe Leu Leu Phe
  90                      95                      100                      105
agc ttt cat gac gct ggt cat tat ctt gct gca tgt att ctg ggg cat      569
Ser Phe His Asp Ala Gly His Tyr Leu Ala Ala Cys Ile Leu Gly His
                      110                      115                      120

tgt att ttt tgatggctgt gagaaga      595
Cys Ile Phe

```

<210> 664
 <211> 487
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 41..226

```

<400> 664
gatacgggttc ctccaccgag gcccatgcga agtttccact atg gct tcc agc act      55
                      Met Ala Ser Ser Thr
                      1                      5
gtc ccg gtg agc gct gct ggc tcg gct aat gaa act ccc gaa ata ccg      103
Val Pro Val Ser Ala Ala Gly Ser Ala Asn Glu Thr Pro Glu Ile Pro
                      10                      15                      20
gac aac gtg gga gat tgg ctt cgg ggc gtc tac cct ttg cca ctg ata      151
Asp Asn Val Gly Asp Trp Leu Arg Gly Val Tyr Pro Leu Pro Leu Ile
                      25                      30                      35
gga atg act tcc gga gaa ctt gat act aaa ttt ggg act ctt tgc tgc      199
Gly Met Thr Ser Gly Glu Leu Asp Thr Lys Phe Gly Thr Leu Cys Cys
                      40                      45                      50
ggg agt ttg gct ggc cag gaa ctt gag tgacattgac ctcatggcac      246
Gly Ser Leu Ala Gly Gln Glu Leu Glu
                      55                      60
ctcagccagg ggtgtagcca agtagacaaa tggaatcctg tgctgaaccc gaatcttcca      306
aaaaacagcc tacaatctgt gaccaccaca agatgtgcc tgaatggcagc tgaagtttga      366
ttcagatggg cacttttctt ccccttccct gcctagtttc cttttgttcc ttgagtcac      426
gcagaattcc attctctggt cagcagacag gcttaagcta aagtattgca tctattctgt      486
a                                                                487

```

<210> 665
 <211> 415
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 54..248

```

<400> 665
acttctgacc tgcgggcccggt agtccgagcc ggggacggcg gcgtcggtgg gtc atg      56
                      Met
                      1
ctc cgg cac tcc ccc tcg ctg tgg gag ctg gtg gag gag cac gtt ccg      104
Leu Arg His Ser Pro Ser Leu Trp Glu Leu Val Glu Glu His Val Pro
                      5                      10                      15
ctc cgg gag cga cgc gaa gtg aag agg att ctg ggg gga ggc ggc ggt      152
Leu Arg Glu Arg Arg Glu Val Lys Arg Ile Leu Gly Gly Gly Gly Gly
                      20                      25                      30
gga cct gag cct gga gct gcg ggc gga ggt ggc gat gtt acg ggc act      200

```

365

Gly Pro Glu Pro Gly Ala Ala Gly Gly Gly Gly Asp Val Thr Gly Thr
 35 40 45
 gct cca aga ggc tgc atc ctc tca agc ccc cag ctc ccg ccc cat ctc 248
 Ala Pro Arg Gly Ser Ile Leu Ser Ser Pro Gln Leu Pro Pro His Leu
 50 55 60 65
 tgacccctct tctcttctgg caccaccgcc tctcctaaag gacctcttgc gccaggagct 308
 ccggcagttg ctccagggtc tccgccacaa agccatctgt gagggcaggg accaggacca 368
 agcttgggtc cagtatagcc ccarggtcc tgcactttgc cttggag 415

<210> 666
 <211> 487
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 13..432

<400> 666
 agtgaggaaa tc atg ggt cag aat cga gag atg ctt ccg ttt tgg atg aac 51
 Met Gly Gln Asn Arg Glu Met Leu Pro Phe Trp Met Asn
 1 5 10
 agc acc ggc agg cgg gaa ggt tgg cag cgt gga tgg cac ggc tac gac 99
 Ser Thr Gly Arg Arg Glu Gly Trp Gln Arg Gly Trp His Gly Tyr Asp
 15 20 25
 aac gag ctc atg gac atg cgg ggc atc ttc ctg gcc ttc gga cct gat 147
 Asn Glu Leu Met Asp Met Arg Gly Ile Phe Leu Ala Phe Gly Pro Asp
 30 35 40 45
 ttc aaa tcc aac ttc aga gct gct cct atc agg tgc gtg gac gtc tac 195
 Phe Lys Ser Asn Phe Arg Ala Ala Pro Ile Arg Ser Val Asp Val Tyr
 50 55 60
 aat gtc atg tgc aat gtg gtg ggc atc acc ccg ctg ccc aac aac gga 243
 Asn Val Met Cys Asn Val Val Gly Ile Thr Pro Leu Pro Asn Asn Gly
 65 70 75
 tcc tgg tcc agg gtg atg tgc atg ctg aag ggc cgc gcg gca ctg ccc 291
 Ser Trp Ser Arg Val Met Cys Met Leu Lys Gly Arg Ala Ala Leu Pro
 80 85 90
 cgc ctg tct ggc cca gcc act gtg ccc tgg cat gat tct tct ctt cct 339
 Arg Leu Ser Gly Pro Ala Thr Val Pro Trp His Asp Ser Ser Leu Pro
 95 100 105
 gct tgc ata act gat cat att gct tgt ctc aga aaa aaa cac cat cag 387
 Ala Cys Ile Thr Asp His Ile Ala Cys Leu Arg Lys Lys His His Gln
 110 115 120 125
 caa agt ggg cct cca aag cca gat gat ttt cat ttt atg tgt gaa 432
 Gln Ser Gly Pro Pro Lys Pro Asp Asp Phe His Phe Met Cys Glu
 130 135 140
 taatagcttc attaacacaa tcaagacat gcacattgta aatacattat tcttg 487

<210> 667
 <211> 476
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 33..218

<400> 667
 gaacaagcgt ccctgatcca gaaggtgttc ag atg gag atg gcg agt tct gct 53
 Met Glu Met Ala Ser Ser Ala
 1 5
 ggc tcc tgg ctc tct ggc tgc ctc atc cct ctc gtc ttc ctc cgg ctg 101
 Gly Ser Trp Leu Ser Gly Cys Leu Ile Pro Leu Val Phe Leu Arg Leu

366

```

      10      15      20
tct gtg cat gtg tca ggc cac gca ggg gat gcc ggc aag ttc cac gtg      149
Ser Val His Val Ser Gly His Ala Gly Asp Ala Gly Lys Phe His Val
      25      30      35
gcc cta cta ggg ggc aca gcc gag ctg ctc tgc cct ctc tcc ctc tgg      197
Ala Leu Leu Gly Gly Thr Ala Glu Leu Leu Cys Pro Leu Ser Leu Trp
      40      45      50      55
ccc ggg acg tac cca agg agg tgaggtggct gcgggtcccca ttcccgcagc      248
Pro Gly Thr Tyr Pro Arg Arg
      60
gctcccaggc tgttcacata ttccgggatg ggaaggacca ggatgaagat ctgatgccgg      308
aatataaggg gaggacgggtg ctagtggagag atgcccaga ggaagtgtc actctgcaga      368
tccttgacgt gcgccttgag gaccaagggt cttaccgatg tctgatccaa gttggaaatc      428
tgagtaaaga ggacaccgtg atcctgcagg ttgcagcccc atctgtgg      476

```

<210> 668

<211> 475

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 90..431

<400> 668

```

gtattatgat gtcgtataat gtaacattct gaaaatacaa catgtccacc taaatttcaa      60
agggctgaaa aaaattttctt gaaagatga atg gac att gtg gcc cat tgt gag      113
                               Met Asp Ile Val Ala His Cys Glu
                               1      5
aaa gcc atg gca gag tca gtt aag aaa cat cag ttt agc act tct gaa      161
Lys Ala Met Ala Glu Ser Val Lys Lys His Gln Phe Ser Thr Ser Glu
      10      15      20
tct tat ctt cca gca act tta ata ctg aac agc ata tcc ttt aat act      209
Ser Tyr Leu Pro Ala Thr Leu Ile Leu Asn Ser Ile Ser Phe Asn Thr
      25      30      35      40
gat ggc tcc ata cag tat aaa gaa aag ctg tat ttc tct gcc tcc gaa      257
Asp Gly Ser Ile Gln Tyr Lys Glu Lys Leu Tyr Phe Ser Ala Ser Glu
      45      50      55
gcc cta gac gct tac att gat gat ttt cac tta aac tat gag cct cct      305
Ala Leu Asp Ala Tyr Ile Asp Asp Phe His Leu Asn Tyr Glu Pro Pro
      60      65      70
gat att gat aca aag gtt aac ctg gat cag agt cct ctt gaa ttt ctt      353
Asp Ile Asp Thr Lys Val Asn Leu Asp Gln Ser Pro Leu Glu Phe Leu
      75      80      85
gct aag tca aat agt ggg gtc cat gga atc gca cca ggc caa agt ggc      401
Ala Lys Ser Asn Ser Gly Val His Gly Ile Ala Pro Gly Gln Ser Gly
      90      95      100
atg gag aag atc ata cag tgg tat gct gca taacaagggtt tctgtcagca      451
Met Glu Lys Ile Ile Gln Trp Tyr Ala Ala
      105      110
acaaattgct tatatgatgg tggc      475

```

<210> 669

<211> 578

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 411..575

<400> 669

```

gagcaagccc tggtggcagc gcagggtcca gtgcagcccc tccccacagc atgctggggg      60

```

367

ctaattctga tgtcatcttt ctgcagaaaa ccattagacc atccctccag actgccaccc 120
 tcaaagccgt ctgcccaggc cccatctgac actcttgaca tctgcaggtc ccagacccta 180
 tgatgtgtcc actctggagg ctccatcatct tcctcggtt gctggccttg cccttggcac 240
 cacacaagca gccttggcct ggcttgcccc aagcccacag agacaacaaa tccaccctgg 300
 caagaagtcc ttcgataaca acatcgtaaa gatgtgtgca catatgagca tcgttgtgga 360
 gttctggctg gagaaagacg agtttggccg gaggatctgg tgataggcaa atg cga 416

Met Arg

1

tgc aga ccc agc agt gtc cat gtg gcc atc ctc act gar gct atc cca 464
 Cys Arg Pro Ser Ser Val His Val Ala Ile Leu Thr Glu Ala Ile Pro
 5 10 15

cca aag atg atc agt ttc tct aca amc tca aag aga atc ttc aaa aak 512
 Pro Lys Met Ile Ser Phe Ser Thr Xaa Ser Lys Arg Ile Phe Lys Xaa
 20 25 30

tct cca aam atg kta aaa agt cag gta tgt cct ctg atc ggt gaa atc 560
 Ser Pro Xaa Met Xaa Lys Ser Gln Val Cys Pro Leu Ile Gly Glu Ile
 35 40 45 50

tcg ggc agc tgg atg tga 578
 Ser Gly Ser Trp Met
 55

<210> 670

<211> 467

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 62..430

<400> 670

atttccggtt ccggtgtcag ttcgaggcgc cgccgccgcc gccgcagccg ccggagccgc 60
 a atg cct aaa gga gga aga aag gga ggc cac aaa ggc cgg gcg agg cag 109
 Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg Gln

1 5 10 15

tat aca agc cct gag gag atc gac gcg cas tgc agg ctg aga agc aga 157
 Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser Arg
 20 25 30

agg cca ggg aag aag agg agc aaa aaa ggt gga gat ggg gct gca ggt 205
 Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala Gly
 35 40 45

gac ccc aaa aag gag aag aaa tct cta gac tca gat gag agt gag gat 253
 Asp Pro Lys Lys Glu Lys Lys Ser Leu Asp Ser Asp Glu Ser Glu Asp
 50 55 60

gaa gaa gat gac tac cag caa aag cgc aaa ggc gtt gaa ggg ctc atc 301
 Glu Glu Asp Asp Tyr Gln Gln Lys Arg Lys Gly Val Glu Gly Leu Ile
 65 70 75 80

gac atc gag aac ccc aac cgg gtg gca cag aca acc aaa aag gtc aca 349
 Asp Ile Glu Asn Pro Asn Arg Val Ala Gln Thr Thr Lys Lys Val Thr
 85 90 95

caa ctg gat ctg gac ggg cca agg agc ttt cga gga gag aac gag aag 397
 Gln Leu Asp Leu Asp Gly Pro Arg Ser Phe Arg Gly Glu Asn Glu Lys
 100 105 110

aga ttg aga agc aga agg caa aag agc gtt aca tgaaaatgca cttggccggg 450
 Arg Leu Arg Ser Arg Arg Gln Lys Ser Val Thr
 115 120

aagacagagc aagccaa 467

<210> 671

<211> 468

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 63..431

<400> 671

```

agaacatcct aatcrwaaaa gttcgaggcg ccgccgccgc cgccgcagcc gccggagccg      60
ca atg cct aaa gga gga aga aag gga ggc cac aaa ggc cgg gcg agg      107
  Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg
    1          5          10          15
cag tat aca agc cct gag gag atc gac gcg cas tgc agg ctg aga agc      155
Gln Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser
          20          25          30
aga agg cca ggg aag aag agg agc aaa aaa ggt gga gat ggg gct gca      203
Arg Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala
          35          40          45
ggt gac ccc aaa aag gag aag aaa tct cta gac tca gat gag agt gag      251
Gly Asp Pro Lys Lys Glu Lys Lys Ser Leu Asp Ser Asp Glu Ser Glu
          50          55          60
gat gaa gaa gat gac tac cag caa aag cgc aaa ggc gtt gaa ggg ctc      299
Asp Glu Glu Asp Asp Tyr Gln Gln Lys Arg Lys Gly Val Glu Gly Leu
          65          70          75
atc gac atc gag aac ccc aac cgg gtg gca cag aca acc aaa aag gtc      347
Ile Asp Ile Glu Asn Pro Asn Arg Val Ala Gln Thr Thr Lys Lys Val
          80          85          90          95
aca caa ctg gat ctg gac ggg cca agg agc ttt cga gga gag aac gag      395
Thr Gln Leu Asp Leu Asp Gly Pro Arg Ser Phe Arg Gly Glu Asn Glu
          100          105          110
aag aga ttg aga agc aga agg caa aag agc gtt aca tgaaaatgca      441
Lys Arg Leu Arg Ser Arg Arg Gln Lys Ser Val Thr
          115          120
cttggccggg aagacagagc aagccaa      468

```

<210> 672

<211> 527

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 85..327

<400> 672

```

cttttctatg tcaattttct ggcaacaaag tctgtggcca agaaagaagc taccgactcc      60
ttcgaccaca aaaagttctt ccaa atg gtc ggc ctg aag aaa aag agt gcg      111
          Met Val Gly Leu Lys Lys Lys Ser Ala
            1          5
gat gat gtg aag aag gtg ttt cac atg ctg gac aag gac aaa agt ggc      159
Asp Asp Val Lys Lys Val Phe His Met Leu Asp Lys Asp Lys Ser Gly
          10          15          20          25
ttc atc gag gag gat gag ctg gga ttc atc cta aaa ggc ttc tcc cca      207
Phe Ile Glu Glu Asp Glu Leu Gly Phe Ile Leu Lys Gly Phe Ser Pro
          30          35          40
gat gcc aga gac ctg tct gct aaa gaa acc aag atg ctg atg gct gct      255
Asp Ala Arg Asp Leu Ser Ala Lys Glu Thr Lys Met Leu Met Ala Ala
          45          50          55
gga gac aaa gat ggg gac ggc aaa att ggg gtt gac gaa ttc tcc act      303
Gly Asp Lys Asp Gly Asp Gly Lys Ile Gly Val Asp Glu Phe Ser Thr
          60          65          70
ctg gtg gct gaa ast aag aag cac tgactgcccc tggtcttcca cctctctgcc      357
Leu Val Ala Glu Xaa Lys Lys His
          75          80
ctgaacaccc aatctcggcc cctcttgcca cctcctgca tttctgttca gttcgtttat      417
gttattttttt actcccccat cccctgtggc cctctaataga caccattctt ctggaaaatg      477

```

369

ctggagaagc aataaagggtt gtaccagtca aaaaaaaaaa aaaatgcgaa

527

<210> 673

<211> 594

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 161..589

<400> 673

```

aaaaggatat acaaagaatg attttgccat tcatacacta aaaaataact ttgaagcaga      60
tgcttatattt gttaaagtta ttgttcttag tggctctgtg tctaatagact gtcctcgtaa    120
gataacacccc tttggagtga atcaacctgg accttatatc atg tat aca act gta      175
                                   Met Tyr Thr Thr Val
                                   1       5
gat gca aat ggg tat ctg aaa aat gga tca gct ggc caa ctc agc caa      223
Asp Ala Asn Gly Tyr Leu Lys Asn Gly Ser Ala Gly Gln Leu Ser Gln
                                   10       15       20
tcc tca cat tta gct ctc caa cta cca tac aac gtg ctt ggt tta ggt      271
Ser Ser His Leu Ala Leu Gln Leu Pro Tyr Asn Val Leu Gly Leu Gly
                                   25       30       35
cgg agc gca aat ttt ctt gac cat ctc tac gtt ggt att ccc cgt cca      319
Arg Ser Ala Asn Phe Leu Asp His Leu Tyr Val Gly Ile Pro Arg Pro
                                   40       45       50
tct gga gaa aaa tct ata cga aaa caa gag tgg act gca atc att cca      367
Ser Gly Glu Lys Ser Ile Arg Lys Gln Glu Trp Thr Ala Ile Ile Pro
                                   55       60       65
aat tcc cag cta att gtc att cca tac cct cac aat gtc cct cga agt      415
Asn Ser Gln Leu Ile Val Ile Pro Tyr Pro His Asn Val Pro Arg Ser
                                   70       75       80       85
tgg agt gcc aaa ctg tat ctt aca cca agt aat att gtt ctg ctt act      463
Trp Ser Ala Lys Leu Tyr Leu Thr Pro Ser Asn Ile Val Leu Leu Thr
                                   90       95       100
gct ata gct ctc atc ggt gtc tgt gtt ttc atc ttg gca ata att ggc      511
Ala Ile Ala Leu Ile Gly Val Cys Val Phe Ile Leu Ala Ile Ile Gly
                                   105      110      115
att tta cat tgg cag gaa aag aaa gca gat gat aga gaa aaa cga caa      559
Ile Leu His Trp Gln Glu Lys Lys Ala Asp Asp Arg Glu Lys Arg Gln
                                   120      125      130
gaa gcc cac cgg ttt cat ttt gat gct atg tgact      594
Glu Ala His Arg Phe His Phe Asp Ala Met
                                   135      140

```

<210> 674

<211> 545

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 297..500

<400> 674

```

atgaggttcc atgcggccaa caagagaggg agatagaaag gaaagggccc cagatgtggc      60
tccccttcca aatactttta gttcccagtt taacctatgt gatcctttca atatttccac    120
cataatgcat ggggtccagtc ctgttctttg ggttgacaaa aataattagc acaagcatac    180
aaaaaggcaa gagtaatatt caaaatactg aagagctatg gtacactaaa gtgccctggg    240
ggtaagtctg tgtgtgccag cctgcagcag gtggtctcag ccctggaagg aacagc atg    299
                                   Met
                                   1
atg gac acc atg gat ccc ttc tgg ttt aat gca ttc aaa agg act aat      347

```

370

Met	Asp	Thr	Met	Asp	Pro	Phe	Trp	Phe	Asn	Ala	Phe	Lys	Arg	Thr	Asn	
			5					10				15				
acc	ata	ctg	cac	cat	ttg	aga	atg	tcc	aag	cac	aca	gat	gca	gca	gaa	395
Thr	Ile	Leu	His	His	Leu	Arg	Met	Ser	Lys	His	Thr	Asp	Ala	Ala	Glu	
		20					25				30					
gag	gtg	cta	ttg	gaa	aaa	aaa	ggg	tgc	acg	gga	gtc	ata	aca	cta	aac	443
Glu	Val	Leu	Leu	Glu	Lys	Lys	Gly	Cys	Thr	Gly	Val	Ile	Thr	Leu	Asn	
		35				40					45					
aga	cca	aag	ttc	ctc	aat	gca	ctg	act	ctt	aat	atg	att	cgg	cag	att	491
Arg	Pro	Lys	Phe	Leu	Asn	Ala	Leu	Thr	Leu	Asn	Met	Ile	Arg	Gln	Ile	
50					55					60					65	
tat	cac	agc	taaagaagtt			ggaacaagat		cctgaaactt		tcctgatcat		tataa				545
Tyr	His	Ser														

<210> 675

<211> 534

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 13..234

<400> 675

tttttttcca	ct	atg	gct	tcc	agc	act	gtc	ccg	gtg	agc	gct	gct	ggc	tcg		51
		Met	Ala	Ser	Ser	Thr	Val	Pro	Val	Ser	Ala	Ala	Gly	Ser		
		1				5				10						
gct	aat	gaa	act	ccc	gaa	ata	ccg	gac	aac	gtg	gga	gat	tgg	ctt	cgg	99
Ala	Asn	Glu	Thr	Pro	Glu	Ile	Pro	Asp	Asn	Val	Gly	Asp	Trp	Leu	Arg	
		15				20				25						
ggc	gtc	tac	cgc	ttt	gcc	act	gat	agg	aat	gac	ttc	cgg	agg	aac	ttg	147
Gly	Val	Tyr	Arg	Phe	Ala	Thr	Asp	Arg	Asn	Asp	Phe	Arg	Arg	Asn	Leu	
30					35				40					45		
ata	cta	aat	ttg	gga	ctc	ttt	gct	gcg	gga	gtt	tgg	ctg	gcc	agg	aac	195
Ile	Leu	Asn	Leu	Gly	Leu	Phe	Ala	Ala	Gly	Val	Trp	Leu	Ala	Arg	Asn	
			50					55						60		
ttg	agt	gac	att	gac	ctc	atg	gca	cct	cag	cca	ggg	gtg	tagccaagta			244
Leu	Ser	Asp	Ile	Asp	Leu	Met	Ala	Pro	Gln	Pro	Gly	Val				
		65						70								
gacaaatgga	atcctgtgct		gaacccgaat		cttccaaaaa		acagcctaca		atctgtgacc							304
accacaagat	gtgccctgat		ggcagctgaa		gtttgattca		gatgggcact		tttcttcccc							364
ttccctgcct	agtttccttt		tgctccttga		gtccacgcag		aattccattc		tctggtcagc							424
agacaggctt	aagctaaagt		attgcctcta		ttctgtaaaag		ttctgtacat		agttcccaag							484
cttctgcagg	gggtgatttt		tgctcttgtc		ctgagaaata		acagtgtctgt									534

<210> 676

<211> 509

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 125..367

<220>

<221> misc_feature

<222> 91,461

<223> n=a, g, c or t

<400> 676

agccgagcct	gcggaaggcg	gcggcggcgg	cacctgcgat	cagcggtctg	ggcagggttat		60
ggtagtgctg	actgcggtgt	gagcagagcg	nccacggggc	ccgccatgcg	ccggcgggccc		120
tgac atg	ggc gcc	acg ggt	cca aag	ctc ggg	gcc tgt	ggc ctt	cgc ctc
							169

371

Met Gly Ala Thr Gly Pro Lys Leu Gly Ala Cys Gly Leu Arg Leu
 1 5 10 15
 ggc ggc cgg ags ggc ggc tca gag gca gca gga gct gag caa gct ttg 217
 Gly Gly Arg Xaa Gly Gly Ser Glu Ala Ala Gly Ala Glu Gln Ala Leu
 20 25 30
 gtr cgg cct cgg ggc cga gct gtr ccc ccc ttc gta ttc acg cgc cgc 265
 Val Arg Pro Arg Gly Arg Ala Val Pro Pro Phe Val Phe Thr Arg Arg
 35 40 45
 ggc tct atg ttc tat gat gag gat ggg gat ctr gct cac gag ttc tat 313
 Gly Ser Met Phe Tyr Asp Glu Asp Gly Asp Leu Ala His Glu Phe Tyr
 50 55 60
 gag gag aca atc gtc acc aag aac ggg cag aak ckg gcc aag ctg agg 361
 Glu Glu Thr Ile Val Thr Lys Asn Gly Gln Xaa Xaa Ala Lys Leu Arg
 65 70 75
 gca tcg tgaagctgga tcasccccgc atccacgtgg atttccctgt gatcctctat 417
 Ala Ser
 80
 gaggtgtgac ccctgggagg tggcagacag aagcaccccc tgcncgggca agaaactccc 477
 aggc tcaatc aagggtgtggc ttccatttga gg 509

<210> 677
 <211> 591
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 101..304

<400> 677
 gcggaagtgc tgcttactgg gtcctgaatc ggctgctgac tgttctaggc atctttcacc 60
 ctgaagtcac gggaacaacg ctgagggggc gccagagatt atg tgg gct ttt cca 115
 Met Trp Ala Phe Pro
 1 5
 tgg gaa gaa cca ctc tac gtt atc acc ttc tac ata gta gca gtt gaa 163
 Trp Glu Glu Pro Leu Tyr Val Ile Thr Phe Tyr Ile Val Ala Val Glu
 10 15 20
 gcc aaa tgg aca gaa agc ccg aga caa cat gaa gtt gtt cta caa gtt 211
 Ala Lys Trp Thr Glu Ser Pro Arg Gln His Glu Val Val Leu Gln Val
 25 30 35
 att ttg gag aaa ttg act tac cat acc act cat caa ccc atg caa aag 259
 Ile Leu Glu Lys Leu Thr Tyr His Thr Thr His Gln Pro Met Gln Lys
 40 45 50
 cct gtc tat gtc caa tca gca gaa tgt ctc gga cca cct aaa aag 304
 Pro Val Tyr Val Gln Ser Ala Glu Cys Leu Gly Pro Pro Lys Lys
 55 60 65
 taaaagaagg agactgaaat aatagcatct ttgatgaaaa ctatctggaa gacaagttgt 364
 taacaattct ggggatcttg gtgattacag agttcttaat ccctctgtcc atagggtgatg 424
 acaattacag gctgcctata ggtcctatag tgctcacaca cctccagccc ttcccatgg 484
 tgtacacaca cttgcagtat attcatctct ttgtcttatt tgagagtagg gctgggtgtg 544
 tgtacaaact aatgacaaat acttgacagt cacacagcag tgataca 591

<210> 678
 <211> 348
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 11..331

<400> 678
 aaaaaacaac atg cca cag tta aaa ctt act tca gtt tgg aca act act 49

372

	Met	Pro	Gln	Leu	Lys	Leu	Thr	Ser	Val	Trp	Thr	Thr	Thr			
	1				5					10						
cac	agc	tac	tac	aca	grg	acc	cga	acg	agk	tca	cct	ggd	tak	ama	cct	97
His	Ser	Tyr	Tyr	Thr	Xaa	Thr	Arg	Thr	Xaa	Ser	Pro	Gly	Xaa	Xaa	Pro	
	15					20				25						
gga	cca	cca	cca	atg	gat	atw	caa	atg	gca	aac	aat	ttt	act	ccg	ccc	145
Gly	Pro	Pro	Pro	Met	Asp	Ile	Gln	Met	Ala	Asn	Asn	Phe	Thr	Pro	Pro	
	30				35				40					45		
tct	gca	act	cct	cag	gga	aat	gac	tgt	gac	ctc	tat	gca	cat	cac	agc	193
Ser	Ala	Thr	Pro	Gln	Gly	Asn	Asp	Cys	Asp	Leu	Tyr	Ala	His	His	Ser	
				50				55					60			
acg	gcc	agt	ata	gta	atg	cct	ctg	cat	tac	agc	ctc	gkc	ttc	atc	att	241
Thr	Ala	Ser	Ile	Val	Met	Pro	Leu	His	Tyr	Ser	Leu	Xaa	Phe	Ile	Ile	
			65				70						75			
ggg	ctc	gtg	gga	aac	tta	cta	gcc	ttg	gtc	gtc	att	ggt	caa	aac	agg	289
Gly	Leu	Val	Gly	Asn	Leu	Leu	Ala	Leu	Val	Val	Ile	Val	Gln	Asn	Arg	
		80					85					90				
aaa	aaa	atc	atc	tca	cct	tgr	gcc	cam	ccc	cam	ccc	att	gkc			331
Lys	Lys	Ile	Ile	Ser	Pro	Xaa	Ala	Xaa	Pro	Xaa	Pro	Ile	Xaa			
	95					100					105					
taatctgtag	aagctaa															348

<210> 679
 <211> 503
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 194..409

<400> 679	
catctcccac	caggccccac
ctgcaacact	ggggatcaaa
tttcaacacg	agattttggaa
60	
ggggcaaata	tccaaactgt
atcagcactc	tttgaaagat
tacagagtaa	atatctttct
120	
tcgtcagaaa	tggaatgata
ccgcctcgc	gtacagttaa
tatcctgacg	actctttaga
180	
cctcgacccc	tcc atg ttg
gac tcc att	tgg aaa cct
gat ttg ttc	ttt
229	
	Met Leu Asp Ser Ile Trp Lys Pro Asp Leu Phe Phe
	1 5 10
gcc aat gaa	aag ggt gcc aac
ttt cat gaa	gtc act aca gac
aac aaa	
277	
Ala Asn Glu	Lys Gly Ala Asn
Phe His Glu	Val Thr Thr Asp
Asn Lys	
	15 20 25
ttg cta aga	att ttc aaa aat
gga aat gtt	ctt tat tca
ata aga	tta
325	
Leu Leu Arg	Ile Phe Lys Asn
Gly Asn Val	Leu Tyr Ser
Ile Arg	Leu
	30 35 40
aca tta aca	ctt tcc tgt
cca atg gat	ctc aag aat
ttt ccc	atg gat
373	
Thr Leu Thr	Leu Ser Cys Pro
Met Asp Leu	Lys Asn Phe
Pro Met Asp	
	45 50 55 60
gta caa aca	tgt ata atg
caa ctg gaa	agc ttg gtt
taaaatgaca	
419	
Val Gln Thr	Cys Ile Met
Gln Leu Glu	Ser Leu Val
	65 70
agataaaaga	ttgtaaatcg
gtggaaccac	tggaagtag
gcaaaaagag	tttaacctac
479	
tggaagcaa	gagctgtcag
aatc	
503	

<210> 680
 <211> 561
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 256..426

<400> 680

373

```

ttgaagattt ttttgggtata ttctagatgt taatctcatc ctgtcttctg tttatcaact      60
ttgtcaatgg taacttttgag aaatttttag ttttgatata gcccagtcoc attaatTTTT 120
gccttaggtt gcactttaag agtataagat atttttcctt accccaagat tacaaagata 180
ttctgtattc tgtcagcttt ccagttttac ctttcacatt gagggcttta ttctatttga 240
agttcatctt catat atg gga aaa caa agg aat ctc aca tta ttt ttt tcc 291
                Met Gly Lys Gln Arg Asn Leu Thr Leu Phe Phe Ser
                  1         5         10
ttt cca ttc tgt gag aca gtt ttc cta aat cat cta tta aac agt gta      339
Phe Pro Phe Cys Glu Thr Val Phe Leu Asn His Leu Leu Asn Ser Val
          15         20         25
ttc ttt cct tca tat agt ttt gtg gag cca ctt tta tca tac ctc aga      387
Phe Phe Pro Ser Tyr Ser Phe Val Glu Pro Leu Leu Ser Tyr Leu Arg
          30         35         40
ttc cca cag aca cat att agt tta ctt ttt caa ctt tat taagagttat      436
Phe Pro Gln Thr His Ile Ser Leu Leu Phe Gln Leu Tyr
          45         50         55
taaggctgag cgtgggtggct caatgcctgt aatcccagca ctttggggaga ctgaggtggg 496
aggattgttt gagtccagga gttcaagacc agcctgggca acatggagag accctgtctc 556
tacaa                                         561

```

<210> 681
 <211> 520
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 21..326

```

<400> 681
ctttggcttc cccaccggca atg gag cgg ttg acg ttg cct ctc ggc ggc gcg      53
                Met Glu Arg Leu Thr Leu Pro Leu Gly Gly Ala
                  1         5         10
gcg gcg gtg gac gag tac ctg gag tac cgg aga att gtt ggt gag gat      101
Ala Ala Val Asp Glu Tyr Leu Glu Tyr Arg Arg Ile Val Gly Glu Asp
          15         20         25
gat gga ggg aaa ctt ttt act cct gaa gaa tat gaa gaa tac aaa aga      149
Asp Gly Gly Lys Leu Phe Thr Pro Glu Glu Tyr Glu Glu Tyr Lys Arg
          30         35         40
aaa gtt tta cct ctg cgc tta caa aac aga tta ttt gtg agc tgg cgg      197
Lys Val Leu Pro Leu Arg Leu Gln Asn Arg Leu Phe Val Ser Trp Arg
          45         50         55
tca cca aca ggg atg gat tgt aaa ctt gtg ggc cca gag aca ctg tgt      245
Ser Pro Thr Gly Met Asp Cys Lys Leu Val Gly Pro Glu Thr Leu Cys
          60         65         70         75
ttt tgt aca cat agg tat aaa caa cat aaa act gac ttg gaa gcg att      293
Phe Cys Thr His Arg Tyr Lys Gln His Lys Thr Asp Leu Glu Ala Ile
          80         85         90
cct cag cag tgc ccc att gat ccc cct gcc aag tgactggctg ccagtgcagg      346
Pro Gln Gln Cys Pro Ile Asp Pro Pro Ala Lys
          95         100
gcttaccttt atgtcccctt gaatggtagc cagcccattc gctgcaggtg caaacacttt 406
gctgatcagc acagtgtctg gcctggcttt acatgcaata catgttccaa gtgttcagga 466
ttccatagct gcttcacttg tgcttgtggt cagcctgcat attggycatg acac      520

```

<210> 682
 <211> 513
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 85..396

<400> 682
actttctcgg ggtgtctgcg taactgcccc gacttgccctt ggtttggtca gatgacctcc 60
tctgggactg gctagccagc gttc atg tct aca tta ttg atc tcg tct tcc 111
Met Ser Thr Leu Leu Ile Ser Ser Ser
1 5
gat gcc tta ttg gag gaa ctg gaa cgc tcc acc ctt cag gac agt gat 159
Asp Ala Leu Leu Glu Glu Leu Glu Arg Ser Thr Leu Gln Asp Ser Asp
10 15 20 25
gaa tat tcc aac cca gct cct ctt ccc ctg gat cag cat tcc aga aag 207
Glu Tyr Ser Asn Pro Ala Pro Leu Pro Leu Asp Gln His Ser Arg Lys
30 35 40
gag act aac ctt gat gag act tcg gag atc ctt tct att cag gat aac 255
Glu Thr Asn Leu Asp Glu Thr Ser Glu Ile Leu Ser Ile Gln Asp Asn
45 50 55
aca agt ccc ttg ccg gcg cas tcg tgt ata cta cca ata tcc agg agc 303
Thr Ser Pro Leu Pro Ala Xaa Ser Cys Ile Leu Pro Ile Ser Arg Ser
60 65 70
tca atg tct aca gtg aag ccc aag agc caa agg aat cac cac cgc ctt 351
Ser Met Ser Thr Val Lys Pro Lys Ser Gln Arg Asn His His Arg Leu
75 80 85
cta aaa cgt cag cag ctg ctc agt tgg atg agc tca ttg ctc acc 396
Leu Lys Arg Gln Gln Leu Leu Ser Trp Met Ser Ser Leu Leu Thr
90 95 100
tgactgagat gcaggccaag gttgcagtga gagcagatgc tggcaagaag cacttaccag 456
acaagcagga tcacaaggcc tccctggact caatgcttgg gggctctcgag caggaat 513

<210> 683
<211> 521
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 317..505

<400> 683
aagggaatct ccgggcccggg tagtgcaggc gccgggtttc ccgcgggtccg agctggcgcg 60
ggcgaggag aatcgctctt aaagggccag cgcacacgag ttcttttggt ccggggccgc 120
aggcggggca ggcccgaactt tcgccgtctt cttgtctact ctccagaacg gccatgattt 180
cccaattctt cattctgtcc tccaaggggg accgctcatc tacaaagact tccgcgggga 240
cagtggcggc cgggatgtgg ccgagctctt ctaccggaag ctgacgggac tgccaggaga 300
cgagtccccg gttgtc atg gac tat ggc tat gta cag acc aca tcc acg gag 352
Met Asp Tyr Gly Tyr Val Gln Thr Thr Ser Thr Glu
1 5 10
atg ctg agg aat ttc atc cag acg gaa gct gtg gtc agc aag ccc ttc 400
Met Leu Arg Asn Phe Ile Gln Thr Glu Ala Val Val Ser Lys Pro Phe
15 20 25
agc ctc ttt gac ctc agc agc gtt ggc ttg ktt ggg gct gag aca caa 448
Ser Leu Phe Asp Leu Ser Ser Val Gly Leu Xaa Gly Ala Glu Thr Gln
30 35 40
cag agc aaa gtg gcc ccc cag cag tgc agc cag ycg scc cgt cct gkc 496
Gln Ser Lys Val Ala Pro Gln Gln Cys Ser Gln Xaa Xaa Arg Pro Xaa
45 50 55 60
cag tcg ctc tgaccagagc caaaag 521
Gln Ser Leu

<210> 684
<211> 462
<212> DNA
<213> Homo sapiens

<220>

<221> CDS

<222> 2..196

<400> 684

```

a atg ctc agc gcg ctg ccc ggc tgg gga ccc gcg cac ctg cag cgc ccg      49
  Met Leu Ser Ala Leu Pro Gly Trp Gly Pro Ala His Leu Gln Arg Pro
    1           5           10          15
ctg ctc ggc cct gca tcc tgc ctg ggc atc ctg cgc ccg gcc atg acg      97
Leu Leu Gly Pro Ala Ser Cys Leu Gly Ile Leu Arg Pro Ala Met Thr
    20          25          30
gcg cac tca ttc gcc ctc ccg gtc atc atc ttc acc acg ttc tgg ggc      145
Ala His Ser Phe Ala Leu Pro Val Ile Ile Phe Thr Thr Phe Trp Gly
    35          40          45
ctc gtc ggc atc gcc ggg ccc tgg ttc gtg ccg aag gac cca acc gcg      193
Leu Val Gly Ile Ala Gly Pro Trp Phe Val Pro Lys Asp Pro Thr Ala
    50          55          60
gag tgatcatcac catgctggtc gccaccgccg tctgctgtta cctcttctgg      246
Glu
65
ctcatcgcca tccctggcgca stgaaccccc tgttcggggc ccagctgaag aatgagacca      306
tctggtacgt gcgcttcctg tgggagtgac ccgccgcccc cgaccaggt gccagctct      366
cggaatgact gtggctccac tgtccctgac aacccttcg tccggaccct ccccccacaca      426
actatgtctg gtcaccagct cctcctgct ggcacc      462

```

<210> 685

<211> 513

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 128..376

<400> 685

```

atctggatat ttctcacatt tgccttggtt tgatgctcct ggagttccca aaagaataga      60
gatcacttta tcaagcttgc agaacaggtt gcatgattgt tgtgctgctg cttgctagta      120
aaagtttt atg ctt gta gaa aga aaa gtt aat ggg ctg gtg ccg cgg ctc      169
  Met Leu Val Glu Arg Lys Val Asn Gly Leu Val Pro Arg Leu
    1           5           10
atg cct gta atc gca gca cat tgg gag gcc cag gtg ggc gca tca ctt      217
Met Pro Val Ile Ala Ala His Trp Glu Ala Gln Val Gly Ala Ser Leu
    15          20          25          30
gag ggt tcc agc atc ttg gac aga gct gtt att gaw cac aat ttg ttg      265
Glu Gly Ser Ser Ile Leu Asp Arg Ala Val Ile Xaa His Asn Leu Leu
    35          40          45
tct gca agc aaa tta tat aat aat att acc ttc gaa gaa ctt gga gct      313
Ser Ala Ser Lys Leu Tyr Asn Asn Ile Thr Phe Glu Glu Leu Gly Ala
    50          55          60
ctt tta gag atc cct gca gct aag gcg gaa aag ata gca tct caa atg      361
Leu Leu Glu Ile Pro Ala Ala Lys Ala Glu Lys Ile Ala Ser Gln Met
    65          70          75
ata acc gaa gac gta tgaatggatt tattgaccag attgatggaa tagttcattt      416
Ile Thr Glu Asp Val
    80
tgaaacacga gaagccctgc caacgtggga taagcagatc caatcacttt gtttccaagt      476
gaataacctt tttggagaaa attagtcaaa cagcacc      513

```

<210> 686

<211> 498

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 118..399

<400> 686

```

aacaattcat gaagttgaag aaaagacact gtcagaaatg aacacagaag cggasaacag      60
cttctccatc acgccagaaa tggcaatgct gaagaagtaa gacaactatt agagacc      117
atg gcg agt aat gaa gtg att gct gac att aat tgc aaa gga aga agt      165
Met Ala Ser Asn Glu Val Ile Ala Asp Ile Asn Cys Lys Gly Arg Ser
1          5          10          15
aag tct aac ttg ggc tgg aca ccc cta cat ctg gca tgc tat ttt gga      213
Lys Ser Asn Leu Gly Trp Thr Pro Leu His Leu Ala Cys Tyr Phe Gly
          20          25          30
cac aga caa gtg gtc cag gat ctg ttg aag gct ggt gca gaa gtg aat      261
His Arg Gln Val Val Gln Asp Leu Leu Lys Ala Gly Ala Glu Val Asn
          35          40          45
gtg ttg aat gac atg gga gac acg ccg ctt cat cga gct gcc ttt aca      309
Val Leu Asn Asp Met Gly Asp Thr Pro Leu His Arg Ala Ala Phe Thr
          50          55          60
gga cga aag gtg aaa atc att cta tgt tca atg ttt gta agt gag gta      357
Gly Arg Lys Val Lys Ile Ile Leu Cys Ser Met Phe Val Ser Glu Val
65          70          75          80
ttt gga gga gta gtt acc att gtt ttc tct gtt ata acc atc      399
Phe Gly Gly Val Val Thr Ile Val Phe Ser Val Ile Thr Ile
          85          90
tgaccagcaa ccgagaaagc cacacaaaaa aatgtatata ccagcacttt ggggtcaaag      459
ccacaggatc ttttgagtct gacagtgagg tccagtact      498

```

<210> 687

<211> 550

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 82..306

<400> 687

```

agtctcgcca gctgccggtc tttcgggggc tccgtaactt tctatccgtc cgcgtcagcg      60
cttgccaccc tcatctccaa t atg cct ggt ccg acc ccc agt ggc act aac      111
          Met Pro Gly Pro Thr Pro Ser Gly Thr Asn
          1          5          10
gtg gga tcc tca ggg cgc tct ccc agc aaa gca gtg gcc gcc cgg gcg      159
Val Gly Ser Ser Gly Arg Ser Pro Ser Lys Ala Val Ala Ala Arg Ala
          15          20          25
ggg gat cca ctg tcc ggc aga gga aaa atg cca gct gtg gga caa gga      207
Gly Asp Pro Leu Ser Gly Arg Gly Lys Met Pro Ala Val Gly Gln Gly
          30          35          40
gtg cag gcc gca caa cct cgg cag gca ccg ggg gga tgt ggc gat tct      255
Val Gln Ala Ala Gln Pro Arg Gln Ala Pro Gly Gly Cys Gly Asp Ser
          45          50          55
aca cag aag att cac ctg ggc tca aag ttg gcc ctg ttc cag tat tgg      303
Thr Gln Lys Ile His Leu Gly Ser Lys Leu Ala Leu Phe Gln Tyr Trp
          60          65          70
tta tgagtcttct gttcatcgct tctgtattta tgttgacat ttggggcaag      356
Leu
75
tacactcggt cgtagattca gttacatcca tctgtcatct gaagaaggag gaaaaaaccc      416
aacatttctt ggaccaaag tatagtgact atctgttcat gagagaaatt ttctgtaagc      476
ttgctgtttt acaggggatt tatcaataat tgattttgag gaatcagttt ttttctatgg      536
ctaataaact tttt      550

```

<210> 688

<211> 501

<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 8..487

<400> 688
agctcct atg agg ccc ctg ccg ccg gtc ggc gat gtc cgg ctg gag ctg 49
Met Arg Pro Leu Pro Val Gly Asp Val Arg Leu Glu Leu
1 5 10
tcg cct ccg ccg ccg ctg ctg ccg gtg ccg gtt gtg agc ggg tct cca 97
Ser Pro Pro Pro Pro Leu Leu Pro Val Pro Val Val Ser Gly Ser Pro
15 20 25 30
gtc ggc tcc tct ggg cgt ctc atg gcc tct agc agc tcc ctg gtg ccc 145
Val Gly Ser Ser Gly Arg Leu Met Ala Ser Ser Ser Ser Leu Val Pro
35 40 45
gac cgg ctg cgc ctg ccg ctc tgc ttc ctg ggt gtc ttt gtc tgc tat 193
Asp Arg Leu Arg Leu Pro Leu Cys Phe Leu Gly Val Phe Val Cys Tyr
50 55 60
ttt tac tat ggg atc ctg cag gaa aag ata aca aga gga aag tat ggg 241
Phe Tyr Tyr Gly Ile Leu Gln Glu Lys Ile Thr Arg Gly Lys Tyr Gly
65 70 75
gaa gga gcc aag cag gag acg ttc acc ttt gcc tta act ttg gtc ttc 289
Glu Gly Ala Lys Gln Glu Thr Phe Thr Phe Ala Leu Thr Leu Val Phe
80 85 90
att caa tgk gtg atc aat gct gtg ttt gcc aag atc tgr tgg gat cgt 337
Ile Gln Xaa Val Ile Asn Ala Val Phe Ala Lys Ile Xaa Trp Asp Arg
95 100 105 110
acc cgg agc tgg ctc tat gct gcc tgt tct atc tcc tat ctg ggt gcc 385
Thr Arg Ser Trp Leu Tyr Ala Ala Cys Ser Ile Ser Tyr Leu Gly Ala
115 120 125
atg gtc tcc agc aat tca gca cta cag ttt gtc aac tac cca act cag 433
Met Val Ser Ser Asn Ser Ala Leu Gln Phe Val Asn Tyr Pro Thr Gln
130 135 140
gtc ctt ggt aaa tcc tgc aag cca atc cca gtc atg ctc ctt ggg gtg 481
Val Leu Gly Lys Ser Cys Lys Pro Ile Pro Val Met Leu Leu Gly Val
145 150 155
acc tct tgaagaagaa gtac 501
Thr Ser
160

<210> 689
<211> 494
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> 282..461

<400> 689
attttcttaa tttttcagaa ctttaaacag atatgtaatt taaaagatgc agccctcct 60
gtgcatgcc tctcatgct gccctctgt tccagggccc acctttcctg tcctttgagg 120
aggaccgcg gcagcaggca ctggagctgg tgggtgcagga ggtctccagc gtgctcaggt 180
cgcgcgagga gaaggtggtg gtctgtcacg tgggtggcca ggctattcag ttcttggtca 240
gcgagacgcc tgccttgggt gctggatgcc ggctgtacga c atg cgg ctc tcc cct 296
Met Arg Leu Ser Pro
1 5
ttc cat tta cag ctg gag ttc cac atg aag gag aag aga gag gac ctc 344
Phe His Leu Gln Leu Glu Phe His Met Lys Glu Lys Arg Glu Asp Leu
10 15 20
cag att agc tgg tct ttc atc agt gtg ccg gaa atg gcc gtt aat atc 392

378

Gln Ile Ser Trp Ser Phe Ile Ser Val Pro Glu Met Ala Val Asn Ile
 25 30 35
 cag ccc aaa cac tgg ggg agg acc agg tgg ctg aga caa gtg cga tgt 440
 Gln Pro Lys His Trp Gly Arg Thr Arg Trp Leu Arg Gln Val Arg Cys
 40 45 50
 ctg acg ttc tca agg aca tct tgaagcattt ggctggttct gcctctccat cag 494
 Leu Thr Phe Ser Arg Thr Ser
 55 60

<210> 690
 <211> 582
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 120..530

<400> 690
 aaagtcgtgc tgcaggcgtc gcgccaatct tcgctctgag gtgctgtctc accggtgaga 60
 cctggaagcg ggcgagtctc gtgctgtgtc ggacctgcag cccctggcct tccgccacc 119
 atg gag tac ctc atc ggt atc caa ggc ccc gac tat gtt ctt gtc gcc 167
 Met Glu Tyr Leu Ile Gly Ile Gln Gly Pro Asp Tyr Val Leu Val Ala
 1 5 10 15
 tcc gac cgg gtg gcc gcc agc aat att gtc cag atg aag gac gat cat 215
 Ser Asp Arg Val Ala Ala Ser Asn Ile Val Gln Met Lys Asp Asp His
 20 25 30
 gag aag atg ttt aag atg agt gaa aag ata tta ctc ctg tgt gtt gga 263
 Asp Lys Met Phe Lys Met Ser Glu Lys Ile Leu Leu Leu Cys Val Gly
 35 40 45
 gag gct gga gac act gta cag ttt gca gaa tat att cag aaa aac gtg 311
 Glu Ala Gly Asp Thr Val Gln Phe Ala Glu Tyr Ile Gln Lys Asn Val
 50 55 60
 caa ctt tat aag atg cga aat gga tat gaa ttg tct ccc acg gca gca 359
 Gln Leu Tyr Lys Met Arg Asn Gly Tyr Glu Leu Ser Pro Thr Ala Ala
 65 70 75 80
 gct aac ttc aca cgc cga aac ctg gct gac tgt ctt cgg agt cgg acc 407
 Ala Asn Phe Thr Arg Arg Asn Leu Ala Asp Cys Leu Arg Ser Arg Thr
 85 90 95
 cca tat cat gtg aac ctc ctc ctg gct ggc tat gat gag cat gaa ggg 455
 Pro Tyr His Val Asn Leu Leu Leu Ala Gly Tyr Asp Glu His Glu Gly
 100 105 110
 cca gcg ctg tat tac atg gac tac ctg gca gcc ttg gcc aag gcc ctt 503
 Pro Ala Leu Tyr Tyr Met Asp Tyr Leu Ala Ala Leu Lys Ala Leu
 115 120 125
 ttg cag ccc acg gct atg gtg cct tcc tgactctcag tatectcgac 550
 Leu Gln Pro Thr Ala Met Val Pro Ser
 130 135
 cgatactaca caccgactat ctcacgtgar ag 582

<210> 691
 <211> 506
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> 4..168

<400> 691
 aag atg gcg gac cgt ggc ggc gtg ggt gaa gcc gca gct gtt gga gcg 48
 Met Ala Asp Arg Gly Gly Val Gly Glu Ala Ala Ala Val Gly Ala
 1 5 10 15

379

tct cct gca tct gtc cct ggc cta aac ccg acg cta ggc tgg agg gag	96
Ser Pro Ala Ser Val Pro Gly Leu Asn Pro Thr Leu Gly Trp Arg Glu	
20 25 30	
cga ctg cgt ccg ggc tgg cgg gga ctg ggg cct cgt tgt ggt tcg tgg	144
Arg Leu Arg Pro Gly Trp Arg Gly Leu Gly Pro Arg Cys Gly Ser Trp	
35 40 45	
cgg ggc tgg ggc tgc ttt acg ccc tgaggatccc tttgaggctg tgtgagaatt	198
Arg Gly Trp Gly Cys Phe Thr Pro	
50 55	
tggcagcggg gactgtatttt ttaaattcoat tgacacccaa attctatgtg gcacttacag	258
ggacctcttc attgatataca ggactaatat ttatatattga atggtggtac ttccataagc	318
atggcacatc ttttattgag caagtatctg taagccattt gcaaccactg atgggaggaa	378
cagagagcag catttcagaa ccaggttctc cttcgaggaa cagagaaaat gaaaccagca	438
gacagaattt gtcaggtgac tactttttcta atgtgttttc agagctgtgt atttaagatt	498
gagtttgg	506

<210> 692

<211> 499

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 64..459

<400> 692

gtcctctggg ccacgctcca cttgcccgc tccacccgga aagcccccca ggctgagtgc	60
ggc atg atc tcc atc acc gaa tgg cag aag att ggt gtg ggg atc acc	108
Met Ile Ser Ile Thr Glu Trp Gln Lys Ile Gly Val Gly Ile Thr	
1 5 10 15	
ggt ttc ggc atc ttc ttc atc ctc ttt gga aca ctc ctg tac ttt gat	156
Gly Phe Gly Ile Phe Phe Ile Leu Phe Gly Thr Leu Leu Tyr Phe Asp	
20 25 30	
tcc gtg ctc ctg gcc ttt gga aac ctg ctg ttc ctg acg ggc ctg tcc	204
Ser Val Leu Leu Ala Phe Gly Asn Leu Leu Phe Leu Thr Gly Leu Ser	
35 40 45	
ctc atc att ggc ctg agg aag acc ttt tgg ttc ttc ttc caa cgg cac	252
Leu Ile Ile Gly Leu Arg Lys Thr Phe Trp Phe Phe Phe Gln Arg His	
50 55 60	
aaa ctc aag gga acc agc ttc ctc ctg ggg ggt gtg gtt atc gtg ctc	300
Lys Leu Lys Gly Thr Ser Phe Leu Leu Gly Gly Val Val Ile Val Leu	
65 70 75	
cta cgc tgg ccc ctc ctc ggc atg ttc ctg gaa acc tac gga ttc ttc	348
Leu Arg Trp Pro Leu Leu Gly Met Phe Leu Glu Thr Tyr Gly Phe Phe	
80 85 90 95	
agc ctc ttt aag ggc ttt ttc cct gtc gcc ttc ggc ttc ctg ggc aat	396
Ser Leu Phe Lys Gly Phe Phe Pro Val Ala Phe Gly Phe Leu Gly Asn	
100 105 110	
gtc tgc aac atc ccc ttc ctg ggt gcg ctg ttc cgg aga ctt caa ggc	444
Val Cys Asn Ile Pro Phe Leu Gly Ala Leu Phe Arg Arg Leu Gln Gly	
115 120 125	
act agc tcg atg gtc tgaaaacaga gatgagctcc ttgaacttgg atcattgggt	499
Thr Ser Ser Met Val	
130	

<210> 693

<211> 756

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 188..616

<400> 693

```

attttattac ccggagtcct ggttcctggg ccagaccggt gtgtcggtttt tggcccaagc      60
tagagaatgt taagggcttc tgcgggtgggt tgggtgctaga ggcgcgcgca acaggtgctg      120
cgggggcggc gcgggagggc gtgcccttgc ttccggatcc ggtctcagct ctgggagggg      180
acgggag atg ttg cag gcg ccg aga ggg cgg gcc agg gcc gca ctc cgg      229
      Met Leu Gln Ala Pro Arg Gly Arg Ala Arg Ala Ala Leu Arg
      1          5          10
aga ctc gcg gtt gct acg cgc acc atg gct gga gcg ccc acg gtc tcg      277
Arg Leu Ala Val Ala Thr Arg Thr Met Ala Gly Ala Pro Thr Val Ser
15          20          25          30
ctt cct gaa ctc cgt tca ctc cta gcc tcc gga cgg gcc cgg ctc ttc      325
Leu Pro Glu Leu Arg Ser Leu Leu Ala Ser Gly Arg Ala Arg Leu Phe
      35          40          45
gac gtg cgc tct cgc gag gag gcg gca gct ggg acc atc cca ggg gcg      373
Asp Val Arg Ser Arg Glu Glu Ala Ala Ala Gly Thr Ile Pro Gly Ala
      50          55          60
ctc aac atc ccg gtg tcc gag ttg gag agt gct ctg cag atg gag cca      421
Leu Asn Ile Pro Val Ser Glu Leu Glu Ser Ala Leu Gln Met Glu Pro
      65          70          75
gct gcc ttc cag gct tta tat tct gct gag aag cca aag ctg gaa gat      469
Ala Ala Phe Gln Ala Leu Tyr Ser Ala Glu Lys Pro Lys Leu Glu Asp
      80          85          90
gag cat ctc gtt ttc ttc tgt cag atg ggc aag cgg ggc ctc cag gca      517
Glu His Leu Val Phe Phe Cys Gln Met Gly Lys Arg Gly Leu Gln Ala
      95          100          105          110
cgc agc tgg ccc gga gtc ttg gat aca ctg ggg ctc gca act acg ctg      565
Arg Ser Trp Pro Gly Val Leu Asp Thr Leu Gly Leu Ala Thr Thr Leu
      115          120          125
gag cct ata gag aat ggt tgg aga aag aga gtt agg cag gag gca gct      613
Glu Pro Ile Glu Asn Gly Trp Arg Lys Arg Val Arg Gln Glu Ala Ala
      130          135          140
tac tgattgccac cccctggccc cttaatggcc accttaacta aggggtgtgaa      666
Tyr
cgggctgact tgggtgaattg ggcaactcct tatagtgttg tgcacacaaa agcatcaa      726
aagaacatt taatcaaaaa aaaaaaaaaaw      756

```

<210> 694

<211> 563

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 296..463

<400> 694

```

aatgtttaat gcccgcacct cgtccccggc caccctgtga ggcagccaga cactctccaa      60
aaagcagaga cagcaggaag aggggagtggt aggcagccca ttcacctggg gaaatgactg      120
ggttgatgat ggacgggtggc ggcacccaag ggggacgtgg acccggttcta ctatgactat      180
gagaccgttc gcaatggggg cctgatcttc gctggactgg ccttcacgtg ggggctcctc      240
atcctcctca gcagaagatt ccgctgtggg ggcaataaga agcgcaggca aatca atg      298
      Met
      1
aag atg agc cgt aac agc agc ctc ggc ggt gcc acc cac tgc act ggg      346
Lys Met Ser Arg Asn Ser Ser Leu Gly Gly Ala Thr His Cys Thr Gly
      5          10          15
gcc agc tgg gaa gcc aag cat ggc cct gcc tct ggc gcc tcc cct tct      394
Ala Ser Trp Glu Ala Lys His Gly Pro Ala Ser Gly Ala Ser Pro Ser
      20          25          30
tcc ctg ggc ttt aga cct ttg tcc ccg tca ctg cca gcg ctt ggg ctg      442
Ser Leu Gly Phe Arg Pro Leu Ser Pro Ser Leu Pro Ala Leu Gly Leu
      35          40          45

```

381

aag gaa gct cca gac tca atg tgacccccag gtggcatcgc aactcctgcc 493
 Lys Glu Ala Pro Asp Ser Met
 50 55
 tcgtgccacc tcatgtttat aataaagccg gcgtcagaga ccgctgcttc ccycacaaaaa 553
 aaaaaaaaaa 563

<210> 695
 <211> 502
 <212> DNA
 <213> Homo sapiens

<400> 695
 agaggcccta caccacccga ggcattggggc tccctgggct gttctgcttg gccgtgctgg 60
 ctgccagcag cttctccaag gcacgggagg aagaaattac ccctgtggtc tccattgcct 120
 acaaagtcct ggaagttttc cccaaaggcc gctgggtgct cataacctgc tgtgcacccc 180
 agccaccacc gcccatcacc tattccctct gtggaaccaa gaacatcaag gtggccaaga 240
 aggtggtgaa gacccacgag ccggcctcct tcaacctcaa cgtcacactc aagtccagtc 300
 cagacctgct cacctacttc tgcyggggcgt cctccacctc aggtgcccac gtggacagtg 360
 ccaggctaca gatgcactgg gagctgtggt ccaagccagt gtctgagctg cgggccaact 420
 tcactctgca ggacagagg gcaggcccca ggttgagat gatctgccag gcgtcctcgg 480
 gcagcccacc tatcaccaac ag 502

<210> 696
 <211> 589
 <212> DNA
 <213> Homo sapiens

<400> 696
 gtccctcctc ctagttgcag tctggtagtt gtcgctggcc gtgtgacggc tcgctgttgc 60
 cctgaaggca ggagagccag ctgccagga aaggtggaaa gtggtagaag ctgacccctg 120
 agccctggca ggtctttaag tgcgtttgtg cagccgattt caaggctaag agagaaagac 180
 tgcctctgat ccctgaagga agaaaaaaaa aaaaaaamcg ggaaaaaac tcaacwkga 240
 aaatgtcccc arggaaaaca aakttgtgga raaggcccca gtgcaraatg aagcccccg 300
 tttaggaggt ggtgaatacc aggagcctgg aggaaatgtt aaaggggttt gggctcmacy 360
 tgccccgggt ttkggagagg atgtgcccaa taggcttgct gataacattg atatgataga 420
 tggagatgga gatgatattg aacggttcat ggaggagatg agagagctaa ggmggaaaat 480
 tagggaactt cagttgaggt actctgcgca ttcttatagg ggaccctcct caccatgatc 540
 atcatgatga gttttgcctt atgccttgaa tcttgagggt aataatcat 589

<210> 697
 <211> 420
 <212> DNA
 <213> Homo sapiens

<400> 697
 gattcatagg agcatataat ataactcttg taaagtagag ggacaattca gtatttttga 60
 caatatgtag aataacatgt tctcttagta ggcataggat atactacatt aagaaatctg 120
 ggctaagcgt ggtgggttcat gcctgtaatc ccagcacttt gggaggccga ggcgggcgga 180
 tcaattgggg tcaggagttc gggaccggcc tggccgacat ggtgaagccc cgtctttact 240
 agaaatacaa aagttggccg ggcgtggtgg tgggcgcctg tggctcagc tactcggggg 300
 gctgaggcgg gagagtcact tgaacccggg aggcggaggt tgcagtgagc cgagatcacg 360
 ccactgcact ccagcctgga tgacagagtg agactctgtc tcaaaaaaaaa aaaaaaaaaa 420

<210> 698
 <211> 496
 <212> DNA
 <213> Homo sapiens

<400> 698
 gacataatca gagctatgct ggaggagaag agggcagcca tttgctggct ggcttgacgt 60
 gagccaggag gtggcaggac gagttaggag gctggttcag tagctcgggc aagagcaggg 120
 cccccagga tctgaaggcc tcccaggccc ccaggcccag cgggtcccag aggagagcga 180
 ggaccccaag gtaactccgg tgagaagggc gaccaggat ttcaggcca gccaggcttt 240

382

ccgggccccac	cggggtccccc	tggattccca	ggcaaagttg	gatcacctgg	cccacctggc	300
cctcaagcag	agaagggcag	cgaagggatt	cgaggcccat	caggcctgcc	tggctccct	360
gggccaccgg	gacctcctgg	gattcagggc	cccgcgggtc	tggatggttt	ggatgggaag	420
gatggcaagc	ctggcttgag	gggggacctg	gtcctgctgg	ccccctgga	ctcatgggac	480
caccgggctt	tagggg					496

<210> 699

<211> 729

<212> DNA

<213> Homo sapiens

<400> 699

atctctgagc	ctccgcgggc	ttctctagca	aatgggtgga	gccggtaccc	ggctgtaagg	60
caagctggga	tgcgcggagt	agcgggtgga	aaactggast	ggagtctca	ctcggccgct	120
gggtgacttc	ggtgcactag	ggatagtaac	agtaccacct	cgtaggggtg	tagaaagacm	180
aagtgaagta	atattttgtaa	agtgtttaga	acagcgtcta	ctgcatggaa	agtagtattt	240
atttgttaga	gagactccag	gaggtcttgc	ccagggattg	acttctccst	tccstgtctc	300
cascccccas	attcmcgttt	gtgagccaag	ccttgctttt	gcaccggccg	ggctctcttg	360
cctggaaatg	cataccatc	tcaaacctgg	gctttaaaat	cacctcctct	gggacgcctt	420
cctctacttc	tcacagttca	ggcctgttct	gccccacaca	gcagaaaatg	cccacgatgc	480
aataccagag	tctgcatggg	accatttcaa	caacccttag	aagctgactc	agctgtctcc	540
tggccctgac	ttttctaaga	tgagaggagc	tgataggaac	aatttgggaa	ccctcaaagg	600
ctgagagtat	cccaccact	tcacagatga	agcaaactga	ggcccagaga	agggaaatta	660
cttgcccaag	atcaccacgc	aagtaagaaa	cagagctgga	gatgagctca	ggccagcacg	720
gkaccacgc						729

<210> 700

<211> 462

<212> DNA

<213> Homo sapiens

<400> 700

caggaaaatc	gcgaggggacg	ggcggccccc	cgacagtcca	gctcctttcg	gctcttgca	60
gaagccctgg	aggctgagga	gagaggtggc	acgccagcct	tcttgcccag	ctcactgagc	120
ccccagtcct	ccctgcccgc	ctccagggcc	ctggccaccc	ctcccaagct	ccacacttgt	180
gagaagtgca	gtaccagcat	cgcgaaccag	gctgtgcgca	tccaggaggg	ccggtaccgc	240
caccccggct	gctacacctg	tgccgactgt	gggctgaacc	tgaagatgcg	cgggcacttc	300
tgggtgggtg	acgagctgta	ctgtgagaag	catgcccgc	agcgctactc	cgcacctgcc	360
accctcagct	ctcgggcctg	agccgccatg	ccctcagcct	gcctcactgc	tgggccaggg	420
tcatgcctat	ataagttggc	atggcaggga	caatggtggg	ca		462

<210> 701

<211> 417

<212> DNA

<213> Homo sapiens

<400> 701

aatgctcagc	gcgctgcccg	gctggggacc	cgcgcacctg	cagcgcccgc	tgctcggccc	60
tgcactctgc	ctgggcatcc	tgcgcccggc	catgacggcg	cactcattcg	ccctcccgggt	120
catcatcttc	accacgttct	ggggcctcgt	cggcatcgcg	ggccctgggt	cgtgccgaag	180
gacccaaccg	cggagtgatc	atcaccatgc	tggtcgccac	cgcggtctgc	tgttacctct	240
tctggctcat	cgccatcctg	gcgcastgaa	ccccctgttc	gggccccagc	tgaagaatga	300
gaccatctgg	trcccagctc	tcggaatgac	tgtggctcca	ctgtccctga	caacccttc	360
gtccggaccc	tccccacac	aactatgtct	ggtcaccagc	tccctcctgc	tggcacc	417

<210> 702

<211> 470

<212> DNA

<213> Homo sapiens

<400> 702

gcgcgcccgc	cgcgtgcccc	gcccgggtag	ctgctgcagc	ctacgctgcc	tccggcactgc	60
ctgccgcgtg	cggagccgga	gcccagacct	gagtggcgcc	gggcccagacg	tggggctcct	120

383

gggccgcggc	ggcggggcggg	cgatgctcca	gaggcctgac	cagccatgga	ggccgaggca	180
ggcggcctgg	aggagctgac	ggacgaggag	atggcggcgc	taggcgaaga	gctagtgcgg	240
cgcctgcggc	gggaggaggc	ggcgcgcctg	gcggcactgg	tcagcgcggc	cgctcatgc	300
aggaggtgaa	tcggcagctg	cagggccacc	tgggcgagat	ccgcgagctc	aagcagctca	360
accggcgtct	rcaggcagag	aaccgtgagc	tgcgcgacct	ctgctgcttc	ctggactcgg	420
agcgccagcg	cgggcggcgc	gccgcacgcc	agtggcagct	cttcggggacc		470

<210> 703

<211> 452

<212> DNA

<213> Homo sapiens

<400> 703

aatgcgcaag	aaggtgcctg	cgggactgga	gcagagsggc	tgcgaggggtc	ttcccagcgc	60
aggcgggttt	tccagtgtta	cttgccgctg	ggcgtggggg	actagctgcc	tttctggcac	120
aggcaggaag	ccgcaaaaag	tttctgagcc	cccgaacctg	tagcggacgt	ggaaaaagaa	180
cgccccctct	caagtgtctg	gctgaaagat	gccaccaggg	gaakggaact	cgggctagct	240
aaggaggcca	ttcttgatgt	tgcttctaga	tctcatgtca	tcaccgagcc	ctcagctgct	300
ggtggcagct	gctcagcaga	cccttgccat	gggaaagaga	cggagtccac	ccaagccat	360
ctgccttcac	ttagctggag	aggtgctggc	tgtggccggg	gactgaagcc	agctgtgctc	420
tatgattgca	actgtgcagg	ggcatcagag	ct			452

<210> 704

<211> 304

<212> DNA

<213> Homo sapiens

<400> 704

atttcgggag	cgggattggt	tcgcgcagga	agcaggetcc	atttttagcgc	cgcccgccgt	60
cgccatctgt	ttcccttccc	tccccatac	ctagcccagag	tctgagccct	aacgagaagg	120
ctgggcctag	gccgctggat	gctggagtga	aaggaaggga	gaaagggaaa	aagcgggaag	180
agtcgagaag	ggagtgttaa	gaggccaagt	gcgacgcgcg	tatccgggca	gacggactga	240
cggacggggc	cgtgcttctg	ccgcggctgc	gcccggggcc	tcttaaccac	tgatctttga	300
aata						304

<210> 705

<211> 269

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 84,216

<223> n=a, g, c or t

<400> 705

agtcgtggac	tcgtgcagct	ggggcgctccg	cagccgctcg	tcaccgcgct	gatgctgttt	60
ctggcgtttg	gcagcccgtg	gggntcgaac	tgctctctctg	cggaaggagg	actgcattgt	120
gtgcggccgc	cgcgctccga	ggtccccggg	cctctgtctc	ccgggcgctc	tccagsagcg	180
ggysttcggg	gmgggtagcc	ggctggagta	cgggmcttc	gggagccgcg	cgccttctcc	240
ggcgtccggg	tcgagcgcag	atccctgtt				269

<210> 706

<211> 476

<212> DNA

<213> Homo sapiens

<400> 706

atttgttcag	cctccataat	tcgaatacca	gggcaggccg	agccagccgt	gcgccgcgct	60
ccagggccca	gggcgccgca	cacgcatcca	cccaccacc	cagcctcgca	gcgccatggg	120
caagaacaag	cagccacgcg	gccagcagag	gcaggggggc	ccgccggccg	cggacgccgc	180
tgggcccgat	gacatggagc	cgaagaaggg	cacggggggc	ccaaggagt	gcggggaggg	240
atcattgtct	gcttagtggc	ctatcttggc	ttgtttatgc	tttgtgtctc	atatcagggt	300

384

gacgaacgga	catgtattca	atthttctatg	aaactgttat	actttctgct	gagtgccttg	360
ggcctgacgg	tctcgtgtgc	tggccgtggc	ctttgccgcc	caccactatt	cgcagctcac	420
acagtttacc	tgtgagacca	cactcgactc	ttgccagtgc	aaactgccct	cctcgg	476

<210> 707

<211> 266

<212> DNA

<213> Homo sapiens

<400> 707

ctctgcgtgc	gccggaggct	gccgtggcgg	gtgggccgcc	tgactttctcc	tcccggccag	60
ttctcgagcg	cctcacggg	cctcgccctg	cagcctcgct	ctcgtggcg	ctgcgcggcc	120
taggggactg	ggctgctggc	ctccgggtgc	gggggtgggg	caggctccga	cctggggcgt	180
cctggcagcg	cgagccgcgg	gatggggggc	cgggccgcgg	aggaggcrcc	gctggygtgt	240
cccttggttg	agagggcgct	gccggy				266

<210> 708

<211> 384

<212> DNA

<213> Homo sapiens

<400> 708

aaaaaacaac	atgccacagt	taaaacttac	ttcagtttg	acaactactc	acagctacta	60
cacagagacc	cgaacgagtc	actgatatac	acctggacca	ccaccaatgg	atatacaaat	120
ggcaacaac	tttactccgc	cctctgcaac	tcctcaggga	aatgactgtg	acctctatgc	180
acatcacagc	acggccagta	tagtaatgcc	tctgcattac	agcctcgtct	tcattcattgg	240
gctcgtggga	aacttactag	ccttggtcgt	cattgttcaa	aacaggaaaa	aatcatctc	300
accttgagcc	cacccccacc	ccattgtcta	atctgtagaa	gctaataaat	aatcatccct	360
ccttgccctag	caaaaaaaaa	aaaa				384

<210> 709

<211> 497

<212> DNA

<213> Homo sapiens

<400> 709

ctctttttgct	ytaacagaca	agccascgay	tttaggctgg	ataatagtca	aattcttacc	60
tcgctctttc	actgctagta	agatcagatt	gcgtttcttt	cagttactct	tcaatcgcca	120
gtttcttgat	ctgcttctaa	aagaagaagt	agagaagata	aatcctgtct	tcaataacctg	180
gaaggaaaaa	caaaataacc	tcaactccgt	tttgaaaaaa	acattccaag	aactttcatc	240
agagatttta	cttagatgat	ttacacaatg	aagaaagtac	atgcactttg	ggcttctgtc	300
cctgctgctt	aatctttgcc	ctgamcctct	taatgctgat	tctgaggaag	atgaagaaca	360
cacaattatc	acagatacgg	agttgccacc	actgaaactt	atgcattcat	tttgtgcatt	420
caaggcggat	gatagcccat	gtaaagcaat	catgaaaaga	tttttcttca	atattttcac	480
tcgacagtgc	gaagaat					497

<210> 710

<211> 471

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 423

<223> n=a, g, c or t

<400> 710

attcccaact	gccagtgatc	tctgaagccg	actctgaggc	tcctcttttg	ctctaacaga	60
cagcagcgac	tttaggctgg	ataatagtca	aattcttacc	tcgctctttc	actgctagta	120
agatcagatt	gcgtttcttt	cagttactct	tcaatcgcca	gtttcttgat	ctgcttctaa	180
aagaagaagt	agagaagata	aatcctgtct	tcaataacctg	gaagaaaaac	aaaataacct	240
caactccgtt	ttgaaaaaaa	cattccaaga	actttcatca	gakattttac	ttagatgatt	300
tacacaatga	agaaagtaca	tgcacttttg	gcttctgtat	gcctgctgct	taatcttgcc	360

385

```

cctgcccctc ttaatgctga ttctgaggaa gatgaagaac acacaattat cacagatacg 420
gantgcccacc actgaaactt atgcattcat tttgtgcatt caagtcggat g 471

```

```

<210> 711
<211> 405
<212> DNA
<213> Homo sapiens

```

```

<400> 711
aatgatacctg cagcagccct tgcagcgagg cccccaggga ggggccagcg cctcccgcgg 60
gccgccttgg gggtgacttg gggcctggac gccagctccc ctctccgagg agctgtgcc 120
atgagcacca agcggcgcoo ggaggaggag caggagcctc tgcgcaagcr rtttctgtct 180
gaggagaaca tggccaccca cttctctcaa ctcagcctgc acaatgacca cccctactgc 240
agccccccca tgaccttctc cccagccctg ccccactca ggagcccttg ctctgagctg 300
cttctctggc gctatcctgg cagcctcatc cctgaggccc tccgtctgct gakgctggg 360
gacaccccca tgttccccct accctgcaac cccagctggg gacat 405

```

```

<210> 712
<211> 160
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> 105
<223> n=a, g, c or t

```

```

<400> 712
aagaaaccca atgatacctg agcagccctt gcagcgagge ccccaggagg gggccagcgc 60
ctcccgcggg ccgccttggg ggtgacttgg ggccwggacg ssagntcccc yctccwagga 120
gctgtgcccc tgagcaccaa gcggcgvtg gaggaggaga 160

```

```

<210> 713
<211> 427
<212> DNA
<213> Homo sapiens

```

```

<400> 713
caatcttcgc tctgaggtgc tgtctcaccg gtgagacctg gaagcgggag agtctcgtgc 60
tgtgtcggac ctgcagcccc tggccttccg ccaccatgga gtacctatc ggtatccaag 120
gccccgacta tgtkctkkkk hvggcytccg accgggtggg ccgccagcaa tattgtccag 180
atgaaggacg atcatgacaa gatgtttaag atgagtgaag agatattact cctgtgtgtt 240
ggakaggctg gagacactgt acagtttgca gaatatattc agaaaaacgt gcaactttat 300
aagatgcgaa atggatatga attgtctccc acggcagsag ctaacttcac acgccgaaac 360
ctggctgact gkcttcggag tcggacccca tatcatgtga acctcctcct ggctggctat 420
gatgagc 427

```

```

<210> 714
<211> 433
<212> DNA
<213> Homo sapiens

```

```

<400> 714
agactgcgtg cagaagggtg ctgtctcagt ggagctgggt catctcagca ggccttggct 60
ccttgaaactt ttggccgcca tgtgcttccc gaagtcctct ctgatgacat gaagaagctg 120
aaggcccgaa tgcaccaggc catagaaaga ttttatgata aaatgcaaaa tgcagaatca 180
ggacgtggac aggtgatgtc gagcctggca gagctggagg acgacttcaa agagggctac 240
ctggagacag tggcggttta ttatgaggag cagcaccag agctcactcc tctacttgaa 300
aaagaaagag atggattacg gtgccgaggc aacagatccc ctgtcccgga tgttgaggat 360
cccgaaccg aggagcctgg ggagagcttt tgtgacaagg tcatgagatg gttccaggcc 420
atgctgcagc ggc 433

```

```

<210> 715

```

<211> 326
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> 126
 <223> n=a, g, c or t

<400> 715
 atggcggtct tggcacctct aattgctctc gtgtattcgg tgcgcgact ttcacgatgk 60
 ctgcgccaac cttactacct tctgtcggcc ctgctctctg ctgccttcct actcgtkagg 120
 aaactnccgc ygcctctgcca cggctctgcc acccaacgcg aagacggtaa cccgtgtgac 180
 tttgactgga gagaaktgga gatcctgatg tttctcagtg ccattgtgat gatgaagaac 240
 cgcagatcca tgcccatga agagggacaa agaatgggat ataggagtgc acagcccagc 300
 ctgacctgtg acatctctgt gtttca 326

<210> 716
 <211> 507
 <212> DNA
 <213> Homo sapiens

<400> 716
 agactgaggc tgaggctggg gaacatcggg cagcatgagc gctgcggctc ttcctgcgca 60
 ccacggctgc ggctcgtgcc tgccggggtc tgggtggtctc taccgcgaac cggcggctac 120
 tgcgcaccag cccgcctgta cgagctttcg ccaaagagct tttcctaggc aaaatcaaga 180
 agaaagaagt tttcccatTT ccagaagtta gccaaagtga acttaatgaa atcaatcagt 240
 tcttgggacc cgtggaaaaa ttcttctactg aagaggtgga ctcccgaaaa attgaccagg 300
 aagggaaaaat cccagatgaa actttggaga aattgaagag cctaggggctt tttgggctgc 360
 aagtccaga agaatatggg ggctggggct tctccaacac catgtactca cgactagggg 420
 agatcatcag catggatggg tccatcactg tgaccctggc agcgcaccag gctattggcc 480
 tcaaggggat catcttggct ggactg 507

<210> 717
 <211> 561
 <212> DNA
 <213> Homo sapiens

<400> 717
 cttcagaatg ctggatgaaa aaattgaaaa gggctcgggat tactgttcgg aagaagagga 60
 tatcacatag caccaatttt accactcaaa ccaggagcta ctactgtgta aatagggtac 120
 accccagttg aaatctttgc aaaggctcgg tctattcagc gaacagcact atagcaaaaag 180
 aagatcgttc catattgtac gccccattaa attacagtgt ttcttaatga acttgcaaag 240
 gaatattgct aaaaacaaac aaaaaaaact gttatctaac tttctttggt gctgctagtt 300
 aaaactttgt gcaacttttc acttctcttg tgtccaggta tgcagcaaaa ttctgcaatt 360
 tcacctttaa gatactgttg gttttacaga tgctctccaa cctattttct ataagatgag 420
 gtagtgggtga actcagataa caaacttctc ttctaaactg gttctgcttc taagacaagc 480
 atctcctgcc ctctctcctt cctccccatc tctcgcacgc agtctagaga tggactgagc 540
 cttgcttctc actggcagtg t 561

<210> 718
 <211> 530
 <212> DNA
 <213> Homo sapiens

<400> 718
 aacatctggg gacagcggga aaacatgagt gactccaagg aaccaagggt gcagcagctg 60
 ggcctcctgg aagaagatcc aacaaccagt ggcacagac tttttccaag agactttcaa 120
 ttccagcaga tacatggcca caagagctct acagggtgtc ttggccatgg cgccctgggtg 180
 ctgcaactcc tctccttcat gctcttggct ggggtccttg tggccatcct tgtccaagtg 240
 tccaagggtcc ccagctccct aagtcaggaa caatccgagc aagacgcaat ctaccagaac 300
 ctgacccagc ttaaagctgc agtgggtgag ctctcagaga aatccaagct gcaggagatc 360
 taccaggagc tgacccagct gaaggctgca gtgggtgagt tgccagagaa atccaagctg 420

387

caggagatct	accaggagct	gacccggctg	aaggctgcag	tgggtgagtt	gccagagaaa	480
tccaagctgc	aggagatcta	ccaggagctg	acccggctga	aggctgcagt		530

<210> 719
 <211> 512
 <212> DNA
 <213> Homo sapiens

<400> 719						
aataatgctc	ccatcaaagt	tcatacaactt	ctcccaagca	gagaaacccg	aaccacacaa	60
ccaggggcag	gatagcctga	agaaacgtct	acaggcaaaa	gtcaaagtta	ttgggggtgca	120
tagcagcctg	gctggaagat	tctgagtgtc	ctgtctgccc	tgggtgggtt	cattctcctg	180
tctgtcaacc	cggctgcatt	aaatcctgcc	tcattgcagt	gtaagttgga	cgaaaaggat	240
ataccaacca	gacttcttct	ttcttatgat	tatcattcac	cttacaccat	ggactgccat	300
agagccaaag	ccagtctggc	tggaactctg	tctctgatgc	tggtttctac	tgtgttgagg	360
ttctgcctag	ctgtgctcac	tgctgtgtctg	cagtggaaac	agactgtctg	acttccctgg	420
ttactcatgt	cctcaaagac	gactcatgay	gctggatatg	aagaactatt	gacttcttgg	480
gaaaaaacgg	agaaatatta	attggaaagt	tt			512

<210> 720
 <211> 507
 <212> DNA
 <213> Homo sapiens

<400> 720						
acctcgccag	ggccggggcc	gggaacgggac	ggaaaactgg	cggcttggtt	tgttcagcgc	60
gtcgcgtcat	ggctctcggt	tcaggcctcc	acggtgtgca	cctggagtca	ggagcgcgct	120
tgtgcgtaca	gttcggttca	gttactaaaa	gatgtaactg	aactgcagat	ccttggtgaa	180
atatctttca	acaaatctct	atatgaggga	ctgaatgcag	agaaccacag	aactaagatc	240
actgtcgtct	tcctgaaaga	tgagaagtac	cattctttgc	ctatcatcat	taaaggcagc	300
gttggtggac	ttctgggtgt	gatcgtgatt	ctgggtcatcc	tggtcaagtg	tggctttttt	360
aaaagaaaat	atcaacaact	gaacttggag	agcatcagga	aggcccagct	gaaatcagag	420
aatctgctcg	aagaagagaa	ttaggacctg	ctatccactg	ggagaggcta	tcagcagtc	480
tgggacttgg	agaccacgca	tcctttg				507

<210> 721
 <211> 408
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> 261
 <223> n=a, g, c or t

<400> 721						
acctttttta	ctttcacagc	aatagtgcag	aatccagaat	ggatgtcctc	tttgtagcca	60
tctttgctgt	gccacttata	ctgggacaag	aatatgagga	tgaagaaaga	ctgggagagg	120
atgaatatta	tcagggtggtc	tattattata	cagtcacccc	cagttatgat	gacttttagtg	180
cagattttcac	cattgattac	tccatatttg	agtcagagga	caggctgaac	agggttgata	240
aggacataac	agaagcaata	nagactacca	ttagtcttga	aacagcacgt	gcagaccatc	300
cgaagcctgt	aactgtgaaa	ccagtaacaa	cgaacactgt	ggaagaaggc	tgotatgact	360
ctttggatgg	gagtctggca	agaggaaatt	ggaagataaa	ataaataa		408

<210> 722
 <211> 371
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> 281
 <223> n=a, g, c or t

<400> 722

agagaaagat	gttccaggca	gaggggaagag	gcctctttcc	tccaggcctg	gagctttcat	60
gacgctggtc	attatcttgc	tgcatgtatt	ctggggcatt	gtattttttg	atggctgtga	120
gaagaaaaag	tggggcatcc	tccttatcgt	tctcctgacc	cacctgctgg	tgtcagccca	180
gaccttcata	agttcttatt	atggaataaa	cctggcgctca	sattttataat	cctgggtgctc	240
atgggcactg	ggcattctta	gctgcgggag	gcactgccga	nagcctgaaa	ctctgcctgc	300
tctgccaaag	caagaacttt	cttctttacr	accagcgctc	cagataacct	cagggaacca	360
gcacttccca	a					371

<210> 723

<211> 404

<212> DNA

<213> Homo sapiens

<400> 723

aagctaggag	ccgcgcragt	cgtagtgctg	ctgtttgctg	gtctccgcgc	gggaccgggg	60
cgcagggggg	cgctgaggcg	agggtgtcat	gtcagacaac	gaggacaatt	ttgatggcga	120
cgactttgat	gatgtggakg	aggatraagg	gctagatgac	ttggagaatg	ccgaagagtt	180
ttagtagtga	cggggttcac	cgtgttagcg	aggatggctc	cgatctcctg	ccttgtgatc	240
cgccgcctc	ggcctcccaa	agtgcctggg	ttacaggaag	gccaggagaa	tgtcgagatc	300
ctccctctg	gggagccgac	cgcagccaac	cagaagcgaa	tcaccacacc	atacatgacc	360
aagtacgagc	gagcccgctg	gctgggcacc	cgagcgctcc	agat		404

<210> 724

<211> 486

<212> DNA

<213> Homo sapiens

<400> 724

aaagtggaag	gcaacgacac	acggcccgtg	gcgcccttcag	ctagggagaa	agagaggacg	60
ggggctggcg	ggcgggatta	gctgccggag	gctgagtttc	cgcagcctga	gtctccatct	120
gtcgctagac	tggagtgcag	tggcgcgac	ttggctcact	tcaacctcca	catcctgggt	180
tcaagcgact	ctcctgcctc	agcctcccaa	gtagctggga	ctacagggtg	gcaccacccat	240
gttcaactaa	tttttghtaat	tttagtgagg	acgaggtttc	accatgttgg	ccaagatggg	300
ctcgatctct	tgaccttggt	atccacccac	cttggcctcc	caaagggtact	gagattahag	360
atgtgagcca	ctgcgsctag	ccagaaaaat	atttattgat	cctaaaagtgg	gagtcagcgc	420
cactgttgmt	gctagtgttg	aatccagaam	catagctgga	catgggatca	gaggatgaac	480
aagaga						486

<210> 725

<211> 374

<212> DNA

<213> Homo sapiens

<400> 725

aaaaaaaaag	tcctagctag	ggtcccaagt	gacagccagc	atcaactata	acatgtaaag	60
gagccagtct	tcagatgatt	ccagcctcag	ccttcaggcc	actcagctgg	tgccaaatag	120
agtagggatg	agctgtcccc	acagagacct	gccagtgca	cattgtgaga	aagccccaga	180
atgttgacag	tcgctctcct	agcccttctc	tgtgcctcag	cctctggcaa	tgccattcag	240
gccaggtctt	cctcctatag	tggagagtaa	ggaagtgggt	gtggaaagcg	attctctcat	300
tctggcaacc	agttggacgg	scccatcacc	gccctscggt	ccgagtcaac	acatactaca	360
tcgtaggtct	tcar					374

<210> 726

<211> 523

<212> DNA

<213> Homo sapiens

<400> 726

aacaagatgg	cggcgctgctg	ggacggctag	cggccctgctg	tggaggcgag	gaatccgcat	60
ctatggagat	gtccctgcat	cccatgactc	ggagctgatg	gccttcatga	cgaggaagtt	120
gtgggacctg	gagcagcagg	tgaaggccca	gactgatgag	atactgtcca	aggatcagaa	180

389

gatagcggcc	ctagaggacc	tggtgcagac	cctccggcca	cacccggccg	aggcaaccct	240
gcagcggcag	gaggaactgg	agacgatgtg	tgtgcagctg	cagcggcagg	tcagggagat	300
ggagcggttc	ctcagtgact	atggcctgca	gtgggtgggc	gagcccatgg	accaggagga	360
ctcagagagc	aagacagtct	cagagcatgg	cgagagggac	tggatgacag	ccaagaagtt	420
ctggaagcca	ggggactcat	tggcgcccc	tgaggtggac	tttgacaggc	tgctggccag	480
cctgcaggat	cttagtgagc	tggtggtaga	gggtgacacc	aaa		523

<210> 727

<211> 492

<212> DNA

<213> Homo sapiens

<400> 727

attttgcgaa	cggcgagcag	cggcgggcggc	gcggagagac	gcagcggagg	ttttcctggt	60
ttcggacccc	agcggccgga	tggtgaaatc	ctccctgcag	cggatcctca	atagccactg	120
cttcgccaga	gagaaggaag	gggataaacc	cagcgccayc	catccacgcc	agccgcacca	180
tgccgctcct	aagcctgcac	agccgcggcg	gcagcagcag	tgagagttcc	agggctctcc	240
tccactgctg	tagtaaccog	ggtccggggc	ctcgggtggtg	ctcctgatgc	ccctcaccca	300
cccctgaaga	tcccagggtg	gcgaggagga	ggagtagggy	bgcctcgggg	ctgggcatcc	360
ggcccctggg	gccacccctt	gtcagccggg	tgggtaggaa	ccgtagactc	gctcatctcg	420
cctgggtttg	tccgcatggt	gtaatcgtgc	aaataaacgc	tcaactccga	tttagcgggtg	480
tatttcttga	ag					492

<210> 728

<211> 467

<212> DNA

<213> Homo sapiens

<400> 728

acagggccac	tgctgtctac	agaagcagtg	aggatgatgc	caggatgatg	tctgcctcgc	60
gcctggctgg	gactctgatg	cccagccatg	gccttctctt	cctgcgtgag	accagaaatc	120
tgggagccct	gcgtggagggt	gtgaaatcca	gacaattgaa	gatggggcat	atcagagcct	180
aagccacctc	tctaccttaa	tattgacagg	aaaccccatc	cagagttag	ccctgggagc	240
cttttctgga	ctatcaagtt	tacagaagct	ggtggctgtg	gagacaaatc	tagcatctct	300
agagaacttc	cccattggac	atctcaaaac	tttgaaagaa	cttaattgtg	ctcacaatct	360
tatccaatct	ttcaaattac	ctgagtattt	ttctaactctg	accaatctag	agcacttgga	420
cctttccagc	aacaagattc	aaagtattta	ttgcacagac	ttgcggg		467

<210> 729

<211> 528

<212> DNA

<213> Homo sapiens

<400> 729

atctttggga	gctgcatttt	tgataaataa	gaaaaacctc	ctcattttta	tgaagtagtt	60
tcaatggagc	aaaaacatgg	aattaaagaa	ccacagttca	actocatcaa	tgaagtaaca	120
tcatttgcca	actaacaac	tgagtctaca	tctctttatt	tgcaatgtaa	aaccccatcc	180
tgtaggggga	aaaaaaaaata	gtactaatgt	tggatttaga	agtccggagt	tcaatttttag	240
gctttggaga	gaagtaatcc	acttttactt	gagatttcca	atctcatatc	accaccatca	300
tcattatcac	cattatcatc	accaccacca	ccaccatgag	caaccaacat	catcatcatc	360
aagaacaact	tgctttttgct	agccaaaagca	atgggaacac	agcactctat	tagggggcaa	420
ggtattaatt	tgcttttttat	ttaattataa	caaaagcgaa	atacatgttt	ttgtaatttc	480
ctatcaataa	tttagaagaa	ttaaataataa	atataaaata	aaaaccaa		528

<210> 730

<211> 379

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 221,233

<223> n=a, g, c or t

```

<400> 730
gtcggcaggt ctgaggcggc gctgtgtgtg tgaagcgtac ctaggcgggg aggcgacatg      60
gagacagggg cggccgagct gtatgaccag gcccttttgg gcatcctgca gcacgtgggc      120
aacgtccagg atttcctgcg cgttctcttt ggcttcctct accgcaagac agacttctat      180
cgcttgctgc gccacccatc ggaccgcatg ggcttycccg nccbkrggcc ggngaagccy      240
tgggksctgc aggtattcaa aacctttgac cacatgrccc cgtcagratg atgagaagag      300
aaggcaggaa cttgaagaga aaatcagaag aaaggaagag gaagaggcca agactgtgtc      360
agctgctgca gctgagaar

```

```

<210> 731
<211> 103
<212> DNA
<213> Homo sapiens

```

```

<400> 731
aaagcaaccg cgtggtacga gtaacaactc caccggctgg cgtcgcgckgc gcttggcsac      60
aagtcacttt tccagatcta ggagcaccat ggacacctcg tca

```

```

<210> 732
<211> 842
<212> DNA
<213> Homo sapiens

```

```

<400> 732
actgtttacc aaacaccttg gtcataataa tgtcattagt ttctccattt ttatttttctg      60
aactgtacat tcacaactta tgtttctttg agattaatag atattggggg aaaaacgcct      120
ttttaggaaa attatagtga aaatttgaca gttgattggc ataatttctt gtttgaatgc      180
tgcctccatt atatagggtcc ttccaggaac tcaaacactg taagtgaaat atgggagtat      240
agtttttatt atttcttctt ttctttttgt tttcataata taatgcagtt tgttcaggaa      300
atcagcacia agcctgatag tactttacta aaatgactgc attctttgga ttccttcagt      360
ctatgggtca agtcactaaa gattcatttt tgttgagtc tttctgccta accccctaata tgtctgttaa      420
aatcttgacg gtttctgccc gtccaatgag cagagctctc tgacttgggt gtatgctgcc      480
aggctgggta ctttcatact ttgttttctt gttttgcttt aaaactacga ctccagcatac      540
attttcccac atacattttt acattgtacc ttaggactca gtcactctca cttaaattga      600
tgacacaagc agctaataac catttctggg tttctgccta accccctaata tgtctgttaa      660
agccaattct ctgggtgtcc agtgagtgtt ggcttttttt ctttccacat tggcacattc      720
acttctccca ctcttgccat gtaagaaata agcatttaca taattggaaa aatctggatt      780
tctgatgcca aagggttaaa gcttcttgga tttcatttca ttgatataca gccactattt      840
ta

```

```

<210> 733
<211> 351
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> 333,345
<223> n=a, g, c or t

```

```

<400> 733
agcaacgggg tttcctgaac cttcggggcca cccgaatcca ccttctcaga cggaaataga      60
atccagggat tcttctaccc atggaaaaac gtgcggagga agcacatgtg ggcgctgggc      120
tgacgtgca gcggcctcct cctcctgagc tgcagcatct gcttgttctg gtgggccaag      180
cgccgggacg tcttgcatat gcccggtttc ctggcgggtc cgtgtgacat gtccaagtcc      240
gtctcgctgc tctccaagca ccgagggacc aagaagacgc cgtccacggg cagcgtgcca      300
gtcgccctgt ccaaagagtc cagggatgtg ganggaggca ccgangggga a

```

```

<210> 734
<211> 200
<212> DNA
<213> Homo sapiens

```

<220>

<221> misc_feature

<222> 43

<223> n=a, g, c or t

<400> 734

tttgcttgac	tgcttccctcc	taaccctcta	sccactagca	ckntacttcc	taaagctggt	60
gtgtcattaa	ctctgttgga	tcaactctct	gggaaaagat	tctgttaatg	taagtgcact	120
tactccctgg	atgttggtcac	tagtctagt	gcttttgcta	aataaacctt	tcttatttct	180
agaaaaaaaa	aaaaaaaaaa					200

<210> 735

<211> 160

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 118

<223> n=a, g, c or t

<400> 735

aaaaggactt	gaattctgat	cgggctgcag	acgctctgca	ccccgccc	gcttctatgg	60
tactttggtg	tcctctctag	cactggcaag	tcgccttcc	cttcgcccgg	cgccagdnct	120
ccctctctat	gacgcgcgag	ttcccggcgc	cgtctcgttk			160

<210> 736

<211> 369

<212> DNA

<213> Homo sapiens

<400> 736

gcactggatg	cacgcagcca	catcagaaag	gcgggttaaga	ggggcggggg	aagcgctctg	60
atccacgatg	tcctgcgagc	tgatcaatca	gcttcacgct	ggaaataacc	atataaaaac	120
ctgaactcaa	agaagagaac	tggaataatc	ctcccctaac	ccacttccca	tggtgagccc	180
cagctgggac	gctgacacat	ggcccacaact	ttttctcctg	cctgatgtct	ggaccacag	240
gggacagatg	gtgggtccctg	gtcactgaag	taccacacaga	acccatgcaa	gaggcacaaa	300
accatctgct	tagtggtcat	gtggcagcag	acagtatctc	agaagtcctc	gcagcatatg	360
acttcactg						369

<210> 737

<211> 382

<212> DNA

<213> Homo sapiens

<400> 737

aaacacacag	atcgctccca	ggctgctggt	cgagaatgag	tattcctata	cgacccagat	60
agactacaac	cccaaggacc	gcctgctcta	tgccctgggac	aatggccacc	aggtcactta	120
ccatgtcatc	tttgccctact	gacacccttg	tccccacaag	cagaagcaca	gaggggtcac	180
tagcaccttg	tgtgtatgtg	tgtgcgcgca	cgtgtgtgta	ggtgggaggc	aggcaacacc	240
aggagcagaa	atgaaagagg	caagaaataa	gtgctatgtg	gcgagaaaaa	aagttttaat	300
gtattggaga	agtttttaaaa	aaccacagaaa	aacgcttttt	ttttttaata	aagaagaaat	360
ttaaaatcaa	aaaaaaaaaa	ar				382

<210> 738

<211> 406

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 197

<223> n=a, g, c or t

<400> 738

agttctagag	gctgggaagt	cggggattga	ggggccagca	tctaccaagg	gccttcttgt	60
tacttcatcc	catgatggaa	ggtggaaggg	taaagagagg	agaagaagca	gtaccgggaa	120
tcttacatca	gtgacaacct	ggacctcgac	atggaccagc	tggaaaaacg	gtcgcggggc	180
agcgggagca	gtgcgngca	gcatgaaaca	caagcgctg	gtscgkcca	tycaacsghm	240
akccacagca	gttcccacac	ctcgggcatt	gaggcagaca	ccaagccccg	ggacacgggg	300
ccggaagaca	gctactccag	cagtgcats	caccgcaagc	tgaaaacctg	cagctcaatg	360
accagtcatg	gcagctccca	cacctcaggg	gtggagartg	gcggca		406

<210> 739

<211> 502

<212> DNA

<213> Homo sapiens

<400> 739

cactttgaaa	gaaaactgcc	atthttgtatt	taatagcctc	caggttcttt	cagtaatgtt	60
atttgcctcg	tgtgtttttg	tgtgtttgtt	gatgtgcgtt	tgtgcatatg	cgtgagtttc	120
attgccgggg	ttggggcaca	attgtggact	ggggccatga	ggcttccctg	gtcccccactg	180
aaccacacct	agttccacat	ttggctgcat	cttgaattat	gcgactcca	gactttctct	240
cctttttttg	ccttggtctt	tgacactcta	aacccttgga	ccatctgaat	ggagcagcca	300
agttcagtc	cacattttctg	tactgttctt	ctttcacagc	tggaatatgt	cacatgatga	360
agttgtatag	aaacagaacc	atggatggat	ggccaggatt	gcggtggtcc	ctagctagat	420
ccccttctta	tcaatcacct	gatagcaaca	gggacagctg	ccaataccct	gctctttact	480
caatggtacc	agggagggag	ca				502

<210> 740

<211> 367

<212> DNA

<213> Homo sapiens

<400> 740

cagcaagccc	aggcaattcg	ggctgtcccc	ggcccccaga	gcccgggcac	cgtggtcac	60
atcgccatcc	tgtctatcct	cctggccgtc	ctcctcacgg	tcctcctggc	tgtgctcata	120
tacacctgct	tcaacagctg	caggagcact	tcctatcag	gcccagagga	ggcaggaggt	180
gtgagaagat	acaccacgca	cctcgcgttc	agcactcctg	ccgagggggc	ttcctgaggg	240
gttcacagag	ggcccacgtg	tcctccacc	tcctcctgg	cccaggctgc	agagcctgag	300
ctgggacacg	ccctgaagct	tctggaccct	gagagagatt	gtttcttttc	tttactatth	360
ttttttt						367

<210> 741

<211> 427

<212> DNA

<213> Homo sapiens

<400> 741

aatgttcttg	tgaatcttag	acaaccccc	caacatcatt	ccctgcttta	tttgcttagt	60
ccattcttgg	tggaatttat	gagatgtcaa	cactcctcat	taggattccc	tccttctgaa	120
gcctttggga	ttttctaccc	accagcaac	caccagaggt	gcagaaacc	ctgcagacat	180
ccttcttgcg	caasgcagtt	gtacctatga	tgcggtcgtc	ctgctgcaga	ccacgcccac	240
caagggcggg	gccacatcca	ctgcagctga	gctcagagcc	ttctgtgggg	tctgcagtgg	300
caaggccagc	cgccccctgt	cttccctctg	gagctgtttg	aggaatgcag	ttcttttgca	360
taagaaaggc	ttttctctcc	agtcattctg	cttgctgata	aagaacactt	aacagcttca	420
ggctttt						427

<210> 742

<211> 408

<212> DNA

<213> Homo sapiens

<400> 742

tcattctatgt	ttatcacatt	ttaataccac	agcacttata	atgatgtcac	tacatataga	60
-------------	------------	------------	------------	------------	------------	----

393

agctcaaaagt	taaggggattt	gctgaagact	gtaaagttta	tggaagaatt	gagacaaaaa	120
tccagtgtag	ctggccactt	atcccagggc	tttttctact	tcatcacaag	gaatgttttg	180
aaagtgtctg	ctttttttat	ccttaaaaatt	cacctgtcag	ggaggcatta	aaaatttgga	240
aatgtatgcc	agcaaaatgt	gagctctgta	ttttttggca	ttcttatgtt	tgggtttaat	300
aagattaaga	aaatgatact	gggaattttc	tttttcctga	aactttgaat	caccctagta	360
agtcaaaagta	ctaaaaaatg	tactagatca	ttaagactta	tgtgctct		408

<210> 743

<211> 449

<212> DNA

<213> Homo sapiens

<400> 743

aacttggttat	ctattgcagg	aaagtgcctg	caatatgcc	ttgtgacaag	gcaccatcaa	60
tcatcatgtc	cctttaccag	gagggcccg	gtcataagcc	actatataaa	gggaatgatg	120
ggtggactgc	tttggacaaa	tggacctgcg	gtaggagaga	gggacaacag	taggagcagg	180
cagatcttgc	tgtttcaacc	aaaacctcat	gctgaccaga	gttgaggaac	agaagaagat	240
ggtgaaggcc	tgcagtcagc	agaatatgtc	aatgttttgg	ttgaagaagc	tgcttgaatc	300
tgggcttttc	tgtgccatgt	gttctcccag	ggccagcaca	aagawgggct	tttgggtgcag	360
gcccaagacc	accataatca	tcattgatta	ttcctctcca	cgccagktct	ctaaataaac	420
tttctcttct	ttctctgaaa	aaaaaaaaar				449

<210> 744

<211> 207

<212> DNA

<213> Homo sapiens

<400> 744

atttccggtt	ccggtgtcag	ttcgaggcgc	cgccgccgcc	gccgcagccg	ccggagccgc	60
aatgcctaaa	ggaggaagaa	agggaggcca	caaaggcccg	gcgaggcagt	atacaagccc	120
tgaggagatc	gacgcgcast	gcaggctgag	aagcagaagg	ccagggaaga	agaggagcaa	180
aaaaaaaaaa	aaaaagctta	tggatac				207

<210> 745

<211> 208

<212> DNA

<213> Homo sapiens

<400> 745

agaacatcct	aatcrwaaaa	gttcgaggcg	ccgccgccgc	cgccgcagcc	gccggagccg	60
caatgcctaa	aggaggaaga	aaggagggcc	acaaaggccg	ggcgaggcag	tatacaagcc	120
ctgaggagat	cgacgcgcas	tgcaggctga	gaagcagaag	gccagggaag	aagaggagca	180
aaaaaaaaaa	aaaaaagctt	atggatac				208

<210> 746

<211> 442

<212> DNA

<213> Homo sapiens

<400> 746

cgtcattttag	gctttttggga	tttgttggta	tggaaaaata	catggattct	catggtactc	60
ggtggttttga	gtagtctggc	gttctctagg	acactgtggg	gaggtaggac	gcagttgtct	120
gcccctgtgc	cccaaacatt	cagctcagga	tggctgtctga	gctkkgkttt	gggaaaactc	180
acaaccctgt	ggatgtaaca	ggagtcagca	gttttgatca	ctggttcatg	gaaacatgtg	240
agaatatata	tgaacatctt	ttgagtcttg	ctggatttgt	cattcattct	ttcaatcact	300
aattcattca	tttatacatt	cagccccctg	tttaaatatg	tgcttcctgc	cggtcagcag	360
gaagtcttaa	gaatggagcc	tgggggtccaa	atgagagccc	tcagtggctg	ttcagggggsc	420
aatccgttat	gtacagtggg	tc				442

<210> 747

<211> 487

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 221

<223> n=a, g, c or t

<400> 747

tctgtggaaa	ggcgaggtct	ggaaagcagc	ccctcccctg	ggagagagcs	cagggcaggc	60
atgctgtcct	gtttccccc	ctcsscaaca	gggctggcst	ccggctctgt	cataccctat	120
ctactcctgc	cttcttccag	ctggcattcc	ctggcagcga	tggcggggaa	actggcagac	180
atgagcaggt	gccatctgta	tccaccacat	actcctgctg	ntgtgggtgtg	gcgaagacaa	240
ggctgcccag	ccccaacacc	acgaggcacc	ctcgggcatg	ctgttgcttc	tgggaccagt	300
gaccagcctg	stgssccctg	tggacccttag	gaggccaagg	caggtgaatt	gcttgagtcc	360
aggagtttga	gaccagcctg	agcaacatgg	tgaaactccg	tctctacaaa	aaacaaaaaa	420
aatccagcgt	ggaggcatgc	acctgtggtt	tcagctactc	aagaagtgag	gcaggaggat	480
tgcttka						487

<210> 748

<211> 601

<212> DNA

<213> Homo sapiens

<400> 748

aaaaagcgcc	attacgcaga	gagaaagtta	cgaggttcgt	ggccgcggtt	tccccaggca	60
gctggcgctg	gaggcttcgg	cgtcacgtgc	tggctctgat	ttttctcgat	gcactgggga	120
aagcggtgga	ctcttatcgt	gggagggctc	ttgatctgtg	atttatagat	aggcacagct	180
actcccgttc	gggaacccaa	cggcagacag	gtcctagtgc	ccatcagata	cccgcggccg	240
ggactcggag	ctgtgggggtg	tggggaggcg	gaggcaccaa	ctaagagcga	cctagcatcg	300
caaagccgcc	ctcggggcct	catggcgggg	cgctcctggg	aaaggcttta	gcgcggtgtc	360
tctctctctg	gccttggttc	tgtgactatc	aggctcctgc	gctgtcgcgg	catccaggcg	420
ttcagaaact	cgtttttcac	ttcttggttt	catcttaata	ccaacgtcat	gtctggttct	480
aatggttcca	aagaaaattc	tcacaataag	gctcggacgt	ctccttacct	agggtcaaaa	540
gttgaacgaa	gccaggttcc	taatgagaaa	gtgggctggc	ttgttgagtg	gcagactata	600
a						601

<210> 749

<211> 210

<212> DNA

<213> Homo sapiens

<400> 749

aaaaattaaa	cctcaccgmc	ggattcaatt	ctctcgttct	caggccgtct	tcctcgccgg	60
agsgtttwct	tacgaaaaga	atgggcttct	ggacattgtt	aaaaggctctg	ctgctctttg	120
caaatgcact	ggcaatcctc	aacgaagacc	gtttccttgt	tccaagagga	tggacactcg	180
gagaactcca	ccaaaccggt	agaagaaaast				210

<210> 750

<211> 380

<212> DNA

<213> Homo sapiens

<400> 750

tgacataaaa	ggttaaagaa	caggcaacac	aatgagcact	taagttttta	acatgtgggg	60
aatagggcat	tttaaaggct	ggaaccagtt	cagaggaaac	aagggttttg	gtagaggtag	120
aaaggtttta	ttaaccctgg	agatctgcaa	atggaggagg	gggtgaagga	ggaatcttaa	180
gacggaggga	gaagagccaa	ggacaaggtc	aggttgaatk	agtwagrggk	gtgtacacah	240
atcctcttgt	gtgctaggta	gtgggcgtgc	atctggatct	gagaggcagt	aaagcagtga	300
tgaagacagt	gcagcagaac	attgaagtcc	tgattgcagc	agtgactctg	caaccttgga	360
taaattattt	caatctttat					380

<210> 751

<211> 245

<212> DNA

<213> Homo sapiens

<400> 751

actctttctga	ggatccggca	agatggcaga	agtagagsag	aagaagaagc	ggaccttccg	60
caagttcacc	taccgcggcg	tggaacctga	ccagctggca	rsatggtggg	sgtctacaac	120
ggcaagacct	tcaaccaggt	ggagatcaag	cccgagatga	tcggccasta	cctgggagag	180
ttctscatca	cctacaagys	cgtaaagcat	ggscggsccg	gcatcggggc	cacccactcc	240
tcccc						245

<210> 752

<211> 369

<212> DNA

<213> Homo sapiens

<400> 752

aagttagtag	actgaaattc	aaagtcattg	tcataactgt	taatgaaagc	agattcaaag	60
caacaccacc	accactgaag	tatttttagt	tatataagat	tggaactacc	aagcatgtgg	120
ctcctgggtc	gtgtaattct	aatctcacgg	atatcctctg	ttgggggata	aggactgtgt	180
ttctttcctt	ttgtggaaaa	tkgtcattct	gaatcttcag	gvcaaacaca	tctggaaggt	240
gatactgtac	aaattatttg	caacacagga	tacagacttc	aaaacaatga	gaacaacatt	300
tcattgtgtg	aacggggctg	gtccactcct	cccaaagca	rgtccactat	ttctgcagaa	360
aatgtttg						369

<210> 753

<211> 491

<212> DNA

<213> Homo sapiens

<400> 753

accaggtgcc	cagtcttcca	gttgcgaggg	caagcaaacc	cgatcatgagc	aactcccttc	60
cccattctctg	ctcaccatgt	ggacgctgaa	atcgctcctg	gtcctgcttc	tgtgcctcac	120
ctgcagctat	gcctttatgt	tctcttctct	gagacagaaa	actagcgaac	cccaggggaa	180
ggtgcmrtry	ggagagcact	ttcggattcg	gcagaaycta	ccagagcaca	cccaaggctg	240
gcttgggagc	aaatggctct	ggcttytktt	tgytgttgtg	ccgtttgtga	tactgmagtg	300
tcaaagagac	agtgagaaga	ataaggagca	gagtcctcct	ggccttcgag	gcttcccat	360
tcgcactcca	ctaaagaaaa	atcaaaatgc	ttctctttac	aaagactgtg	tattcaatac	420
cttaawsgaa	ctygargtg	agcttwtgaa	atttgtgtcc	raagtgcrra	atcttaasg	480
tgccatggca	a					491

<210> 754

<211> 336

<212> DNA

<213> Homo sapiens

<400> 754

ataaatgctt	tccaatcagt	gctcccatc	ttcattctat	tcaccactta	gttactagaa	60
tgtttttctc	aactatgaat	ctgacttcac	tcccattctt	gaaaacctcc	tgtagtccct	120
gttgtgtacg	ggataaagac	tgtgagcccc	tggagttcag	ggaccattac	ttaatcatat	180
tccagaacct	atcacaatga	atgacacgaa	gaatatattt	gataaaaatc	tgatgaatga	240
accactctaa	aatgcaactc	taataaagaa	aattactgcc	tawgtcrgdd	skgkmtcams	300
gktayaaacc	gtdaaaatta	tgtatagtct	gataaa			336

<210> 755

<211> 497

<212> DNA

<213> Homo sapiens

<400> 755

ataaatgctt	tccaatcaat	gctcccatc	ttcattctat	tcaccactta	gttactagaa	60
tgtttttctc	aactatgaat	ctgacttcac	tccwttctt	gaaaacctcc	tgtagtccct	120
gttgtgtacg	ggataaagac	tgtgagcccc	tggagttcag	ggaccattwc	ttaatcatat	180
tccagaacct	atcacaatga	atgacacgaa	gaatatattt	gataaaaatc	tgatgaatga	240
accactctaa	aatgcaactc	taataaagaa	aattactgcc	tattcagttg	gaactgtacc	300

396

aygttacaaa	cgtaaaatta	tgtatagtct	gataaagaaa	tactrtaata	gccttggaaa	360
cagctcacat	acaatTTTTT	ttataaatgg	attcattacc	atataactgg	accagcttac	420
tgaagagtag	ccctacagca	gtcacttcgc	agtttgcaat	tcatttgagg	ttcaaagcta	480
gctcaaggga	aaacaac					497

<210> 756

<211> 400

<212> DNA

<213> Homo sapiens

<400> 756

atcattttcc	ctgcaggaca	cccggcctcg	gtcctgcttt	agtattcatc	ccagcttcat	60
tactgcagcc	acctcgcaac	ccaccctggg	cagctgctaa	aagggggaaa	aattaaagca	120
attcatactc	ctgggtgactt	actgctgcag	aataattggg	ttctgtttgc	aaggaaacatt	180
atcacagcaa	agcccagaaa	acagaaaggg	atgatcatgg	tgattttact	gctgcttgtg	240
gctatcgtgg	ttgttgacgt	ctggccgacc	aactgatggc	agtaaagaga	ccaccagcag	300
tgacacctgc	caatgacaga	tgcaagccca	acaccctttt	ggtacgcaaa	acctgctctc	360
aataaattcc	cccaaagctc	caaaaaaaaa	aaaaaaaaag			400

<210> 757

<211> 524

<212> DNA

<213> Homo sapiens

<400> 757

acttccctcg	tgcccccgcc	gcgggcccag	atggattccg	ggtgctgggt	gttcggcggc	60
gagttcgagg	actcgggtgt	cgaggagagg	ccggagcggc	ggtcaggacc	gccgcgtcct	120
actgcgccaa	gctctgcgag	ccgcagtggg	tttatgaaga	aacagaaagc	agtgatgatg	180
ttgaagtgct	gactctcaag	aaattcaaa	gagacctggc	ctacagacga	caagagtatc	240
agaaagcact	gcaggagtat	tccagtatct	ctgaaaaatt	gtcatcaacc	aattttgccca	300
tgaaaaggga	tgtccaggaa	ggtcaggctc	ggtgtctggc	tcacctgggt	aggcatatgg	360
aggcgctgga	gattgctgca	aacttgga	ataaagcarg	carcacagac	catttaacca	420
csgtactcta	cctccagctt	gctatttgtt	caagtttgca	gaacttgagg	aaaacaattt	480
tctgcctgca	gaaactgatt	tctttgcac	cttttaatcc	ttgg		524

<210> 758

<211> 543

<212> DNA

<213> Homo sapiens

<400> 758

aagaaggag	gtggctgccc	tccgtcgtgg	tctggcgtgt	attccgagcs	ttgggtgtctg	60
gcggtttccg	agcgtttggtg	tctggcgtgt	tccgaccgtt	ggtgtctggc	ggtttccgac	120
cgttgggtgc	tggcacgcgc	caccctctct	tgctttgggt	gcgccatgcc	gatgtaccag	180
acaagaagac	aagaaaatga	tttgaggaca	gcttcaatcg	cgggtgtgaag	aagaaagcag	240
caaaacgacc	actgaaaaca	acgccgggtg	caaaatatcc	aaagaaaggg	tcccaagcgg	300
tacatcgtca	tagccggaaa	cagtcagagc	caccagccaa	tgatmttttc	aatgctgcga	360
aastgccaaa	agtacatgc	agggatgtcc	ttcctgagat	ccgtgctatc	tgcatagagg	420
aaattgggtg	ttggatgcaa	agctacagca	cgtctttcct	caccgacagc	tatttaaaat	480
atattgggtg	gactctgcat	gataagcacc	gagaagtccg	cgtgaagtgc	gtgaaggctc	540
tga						543

<210> 759

<211> 850

<212> DNA

<213> Homo sapiens

<400> 759

aagtttcatg	ggttcctgga	cgctcagcca	ggggtagtct	ctgaagggcc	gcctggcaat	60
tgggagcaga	agccagtttc	ccgccatctg	ctctccgtgg	aggcagtgtc	ctcgcggctg	120
ctgcccggtg	gggtccagct	gaggaaggag	cgcggtatcc	aggctcgttc	tttgctggg	180
ctgccccctg	ggccccctgg	gcgggtgctg	ctgcggcgcc	agctccaagg	gcgkatccag	240
gcgggagggg	cctcctcgga	gaagcggggc	gcgggtccaa	ctacgcagag	gctggcacgc	300

397

cgaccctcca	cacctcacca	cgcccccatc	tccgtccgtg	tacacacact	cacacaagga	360
cgccaacccc	acctagatgc	aaagcaggat	tcaaaagaac	atctttgcgt	tttctaccgg	420
ctccccatca	tctgactagg	gaggaagaag	cgggtgagaa	acaaaacttc	tttccattgt	480
cctgcccgtt	tctgcggaact	tgttctgagg	ccgaggcacc	tctaagatac	tgatggctct	540
gcagaggacc	cattcattgc	ttctgctttt	gctgctgacc	ctgctggggc	tggggctggt	600
ccagcctcct	atggccagga	tggcatgtac	cagcgattcc	tgcggcaaca	cgtgcaccct	660
gaggagacag	gtggcagtga	tgcctactgc	aacttgatga	tgcaaagacg	gaagatgact	720
ttgtatcact	gcaagcgctt	caacaccttc	atccatgaag	atatctggaa	cattcgtagt	780
atctgcagca	ccaccaatat	ccaatgcaag	aacggcaaga	tgaactgcca	tgaggggtga	840
gtgaggtcac						850

<210> 760

<211> 532

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> 524

<223> n=a, g, c or t

<400> 760

ataaaacaaa	atagtatgca	acttacagca	atgtctctcc	cacttataat	aaaacaccta	60
actctcatgt	agcattttat	ccatattgag	agctctgtga	gatagaaatt	attatattac	120
aagtcacaga	tatgtaaaact	gagatgcaga	gacatcaata	acttgctcaa	ggttacaagg	180
tggacttagt	agttttaagc	agaaccatga	aaacctctgt	gacaactccc	tccagctcca	240
agagtgcgg	gaggtggggg	gcggcgcac	cgcggcctcg	agcttgctac	ctcagcccat	300
ccccaccacc	cctgacatcg	agaacgctga	gtcaccccc	atcttgccct	tcctgttcct	360
tggcaatgag	caggatgctc	aggacctgga	caccatgcag	cggtgaaca	tcggctacgt	420
catcaacgtc	accactcatc	ttccccctta	ccactatgag	aaaggcctgt	tcaactacaa	480
gcggctgcca	gcactgacag	caacaagcag	aacctgcggc	agtntttgaa	ga	532

<210> 761

<211> 275

<212> DNA

<213> Homo sapiens

<400> 761

tgtttgtttt	gcttctacta	agactgtttt	gtttcaaaaa	ggaaacaagt	tttgtgtttg	60
ctgtctacgc	tggagtctctg	aactgtgggt	agaaaacacg	acctggcttt	gtagaaagga	120
cacagggctg	ttttatgaac	taagcgggtga	ggctcagggtg	gcggctctcg	cagaccctcg	180
atgctgttgt	tcttkgaggg	cttaaggcct	gatgaacgta	ggcacgtgat	gcataatagt	240
cttcaatggt	acacttaact	agtctcttct	gtgta			275

<210> 762

<211> 546

<212> DNA

<213> Homo sapiens

<400> 762

aggtatctag	gtatcctccc	atctcaggcc	agactggaga	ttgagcgcct	gttgtatacg	60
ggccctgggc	taacctctcc	cccatcacac	aactcaacag	caacgcccac	gccacaactt	120
ctctggacct	caggttccct	acctgcaaaa	cagtttgac	ctcctccctc	cacaaaactg	180
caaccggggc	cggcgcgggtg	gtcacgcct	gtcatcttaa	caccttgga	ggcggagggtg	240
ggcggatcat	cagaggtcag	gagttcgaga	ccagcctggc	caacaagggtg	aaacccccatc	300
tgtactaaaa	atacagaaat	cagccgggca	tgggtggctcg	cgctgtagt	cccagctact	360
tgggaggctg	aggtaggaga	atcgcttgaa	tccgggagggt	ggaagtcgca	gtgagccgag	420
atcgcgccac	tgcactccag	cctgggtgac	acagtgagac	tccacctcaa	aaacaaaaca	480
aaacaaaaaa	ctgcccgcag	agagctgccc	cgaccacaag	gcctgaatcc	agccttccag	540
ccccag						546

<210> 763

<211> 545

<212> DNA

<213> Homo sapiens

<400> 763

ctggtgtctc	tcaaaagcct	gatcctgcc	aaaccaagaa	tgcgcgcaa	aggaagccat	60
ccacttctga	tgattctgac	tctaattttg	agaaaattgt	ttcgaaagca	gtcacaagca	120
agaaatccaa	gggggagagt	gatgacttcc	atatggactt	tgactcagct	gtggctcctc	180
gggcaaaatc	tgtacgggca	aagaaacct	taaagtacct	ggaagagtca	gatgaagatg	240
atctgtttta	aaatgtgagg	cgattatfff	aagtaattat	cttaccaagc	ccaagactgg	300
ttttaaagtt	acctgaagct	cttaacttcc	tccctctgta	atttagtttg	gggaaggtgt	360
ttttagtaca	agacatcaaa	gtgaagtaaa	gcccaagtgt	tctttagctt	tttataatac	420
tgtctaaata	gtgaccatct	catgggcatt	gttttcttct	ctgctttgtc	tgtgttttga	480
gtctgctttc	ttttgtcttt	aaacctgatt	tttaagttct	tctgaactgt	agaatagcta	540
tcttg						545

<210> 764

<211> 500

<212> DNA

<213> Homo sapiens

<400> 764

gacctgggct	gcagttaggt	tgtcgcctgc	aggaggtgtg	ctgttgaggg	agaggaaaac	60
cgcccgagga	gaggatcaag	tcttccatac	cagggaccag	gccaaaacca	gactgccact	120
gcccctactt	gggccccact	gtccccagaa	agcagctgct	gtgatcatgg	gcaatatctt	180
tggaaacctt	ctcaagagcc	tgattgggaa	gaaggagatg	cgcatcctga	tgggtgggct	240
ggatgccgca	gaaagaccac	catcctatac	aagctgaaac	tgggggagat	cgtcaccacc	300
atccctacca	ttgggttcaa	tgtggagaca	gtggagtata	agaacatcag	ctttacagtg	360
tgggatgtgg	gtggccagga	caagattcga	cccctctgga	gacactactt	ccagaacacc	420
caagggttga	tatttgtggt	cgacagcaat	gatcgggagc	gagtaaata	ggcccgggaa	480
gagctgatga	gaatgctggc					500

<210> 765

<211> 527

<212> DNA

<213> Homo sapiens

<400> 765

gctttgacaa	aaccaggga	agagacagag	agagataatg	gcggctaaac	taatctgttc	60
gtctttaaca	gtgcattcaa	tggcaaataa	gaagccatca	ccttctgctg	caacaagaac	120
aataacctca	aagaagagca	cagcgctcca	caggtgaaac	tgctgacaag	agtcgagcag	180
ctcaagcttc	tgaccaaagc	cgagaaagca	ggacttttgt	ctttagcaga	gaaatcaggt	240
ttctctctat	cgaccatcga	gcgtcttgga	ttgctgacca	aagcagagga	gttcggcggt	300
ttgtctgccc	ccacaaaccc	ggaaacgcct	ggaacgttat	tcacttttag	cctcgggtta	360
ctccttcttg	gaccggtttt	tgcataatgt	gttcctgaag	attacacttg	ggaagtagtg	420
attcaggttc	ttgtggctct	actctctgtt	cttgggtggt	ctgctgcttt	tgtgtcttct	480
ggttttgtct	ccaatttgca	gaaatctgat	tagcccgatc	gtttttt		527

<210> 766

<211> 921

<212> DNA

<213> Homo sapiens

<400> 766

agtagggagg	aaaaccggag	gagagcgag	gaggaaacag	taccggctgg	aggccggtct	60
tgaggagagc	ggggactgct	gggggcgggg	cttgggtggt	accgctggcg	gggcggggcc	120
tgcggctcag	aggggtgggc	tttgagatc	agagggtcga	cgctgcttcg	ttgcctggac	180
tctggtttcc	gccctggagc	aagccggggc	ctggctcgga	gctgggcccgc	catggagtcc	240
acgctggggc	cgggcatcgt	gatagccgag	gcgctacaga	accagctagc	ctggctggag	300
aacgtgtggc	tctggatcac	ctttctgggc	gatcccaaga	tcctctttct	gttctacttc	360
cccgcggcct	actacgcctc	ccgccgtgtg	ggcatcgcg	tgctctggat	cagcctcatc	420
accgagtggc	tcaacctcat	cttcaagtgg	tttctttttg	gagacaggcc	cttttgggtg	480
gtccatgagt	ctggttacta	cagccaggct	ccagcccagg	ttcaccagtt	cccctcttct	540
tgtgagactg	gtccaggcag	cccttctgga	cactgcatga	tcacaggagc	agccctctgg	600

399

cccataatga	cagccctgtc	ttcgcaggtg	gccactcggg	cccgcagccg	ctgggtaagg	660
gtgatgccta	gcctggctta	ttgcaccttc	cttttggcgg	ttggcttgte	gcgaatcttc	720
atcttagcac	atttccctca	ccaggtgctg	gctggcctaa	taactggcgc	tgtcctgggc	780
tggctgatga	ctccccgagt	gcctatggag	cgggagctaa	gcttctatgg	gttgactgca	840
ctggccctca	tgctaggcac	cagcctcatc	tattggaccc	tctttacact	gggcctggat	900
ctttcttggt	ccatcagcct	a				921

<210> 767

<211> 570

<212> DNA

<213> Homo sapiens

<400> 767

aaggggaatct	ccgggcgggg	tagtgcaggc	gccgggtttc	ccgcgggtccg	agctggcgcg	60
ggcggaggag	aatcgctctt	aaagggccag	cgcacacgcg	ttcttttggt	ccggggccgc	120
aggcggggca	ggcccgaactt	tgcgcgtctt	cttgtctact	ctccagaacg	gccatgattt	180
cccaattctt	cattctgtcc	tccaaggggg	accgctcatc	tacaaagact	tccgcgggga	240
cagtggcggc	cgggatgtgg	ccgagctctt	ctaccggaag	ctgacgggac	tgccaggaga	300
cgagtccccg	gttgtcatga	ggcagaggtt	gcagtgaacc	gagatccgga	caccgcactc	360
cagcctggga	gacagagcga	gactccctct	cgggcatcac	catggccgtc	atttcattca	420
catcagacac	agcggcctct	atgttggtgt	cacaacttca	gaaaacgttt	ctcccttcag	480
cctcctagaa	ctgctctcca	ggttggccac	ccttctgggc	gattactgtg	gctccctggg	540
cgaggggaca	tctcccgmaa	tgtggctctg				570

<210> 768

<211> 490

<212> DNA

<213> Homo sapiens

<400> 768

cactcgctgc	ctcggcagcg	ctgctcttct	aagatggctg	ccgctaccgg	tgcggtggca	60
gcctcggccg	cctcgggtca	ggcgggaaggt	aaaaagatca	ccgatctgcg	ggatcatcgat	120
ctgaagtccg	agctgaatcg	gcggaactta	gacatcaccg	gagtcaagac	cgtgctcatc	180
tcccgaactca	agcaggctat	tgaagaggaa	ggaggcgatc	cagataatat	tgaattaact	240
gtttcaactg	atactccaaa	caagaaaacca	actaaaggca	aaggtaaaaa	acatgaagca	300
gatgagttga	gtggagatgc	ttctgtggaa	gatgatgctt	ttatcaagga	ctgtgaattg	360
gagaatcaag	aggcacatga	gcaagatgga	aatgatgaac	taaaggactc	tgaagaattt	420
ggtgaaaatg	aagaagaaaa	tgtgcattcc	aaggagtta	tctctgcaga	agaaaacaag	480
agagctcatg						490

<210> 769

<211> 491

<212> DNA

<213> Homo sapiens

<400> 769

cactcgctgc	ctcggcagcg	gctgctcttc	taagatggct	gccgctaccg	gtgcgggtggc	60
agcctcggcc	gcctcgggtc	aggcgggaag	taaaaagatc	accgatctgc	gggtcatcga	120
tctgaagtcc	gagctgaatc	ggcgggaactt	agacatcacc	ggagtcaaga	ccgtgctcat	180
ctcccgaactc	aagcaggcta	ttgaagagga	aggaggcgat	ccagataata	ttgaattaac	240
tgtttcaact	gatactccaa	acaagaaacc	aactaaaggc	aaaggtaaaa	aacatgaagc	300
agatgagttg	agtggagatg	cttctgtgga	agatgatgct	tttatcaagg	actgtgaatt	360
ggagaatcaa	gaggcacatg	agcaagatgg	aatgatgaa	ctaaaggact	ctgaagaatt	420
tggtgaaaat	gaagaagaaa	atgtgcattc	caaggagtta	ctctctgcag	aagaaaacaa	480
gagagctcat	g					491

<210> 770

<211> 403

<212> DNA

<213> Homo sapiens

<400> 770

cgcccacgcg	gaactcgaac	tcctgcaggg	cctcgtagtc	cagcgacctr	agggcgaaca	60
------------	------------	------------	------------	------------	------------	----

400

ggtggccggt	gtctgtgttg	atggagacca	gggaggcgag	gggcagaatc	gcggtcttga	120
atgctgaaaa	gagctgcaac	ggtctcagga	gaatcttcgc	ggatggggct	tgtgaatgaa	180
gccatgggtca	gttcgggagc	gttatcggtg	atatctacca	cctgtattat	gacagtgcac	240
tttccagaaa	gacctgcccc	atcagaggct	ttaatatcca	cctgatatga	cacaatttww	300
wmaaaatcta	gttttctgat	tattttaatt	tctccccttc	gttcattaag	tgcaaatgtc	360
ttagaaatct	cttcatacacc	ataaaaaatg	agtagaataa	aaa		403

<210> 771

<211> 374

<212> DNA

<213> Homo sapiens

<400> 771

attcttttcc	cccttctcac	cggccaggat	gcagacataa	tgatggcagg	agctggagca	60
gccacctgag	gacccagagc	tcaaagccac	atgttgagaa	gggcagagat	aactgtatcc	120
actctggact	gctgaccttt	gaactattat	gttatttcca	gggaaatgca	aaccaaagga	180
tgtggctctc	gatctaatac	ttagagaatg	tgaccatgaa	gacacttttc	ctacctggta	240
aacaaaagat	aatgagaaaa	gtgagggttg	aagttggttt	actgagccag	gagctataac	300
aggtgctgga	gcaggggctg	tgatctgaat	gaccagaggg	aaggactgat	ggaattggat	360
ggtgakagct	ccag					374

<210> 772

<211> 380

<212> DNA

<213> Homo sapiens

<400> 772

agaaacagtc	tctccccagc	tcacttggtc	gaagatacta	aagagaacta	gatgttatct	60
agtgttatcc	tgctgatccc	aatttcctat	agttctggca	ccagaaaatc	ctgaatgata	120
ggtttttctg	aagctgctga	catcgggtca	gccctcaaca	ccatcacctt	tggccatggt	180
caggaagtac	tcaaagatca	tggtcaaaa	gtctgcagag	cagagacaag	agattgacaa	240
agaggtgttg	atgagggcaa	taaaagcaaa	ggtggcttta	tcacgggtgag	ctacttctca	300
caggagtcgg	gatccacatc	tgacagactac	acaaaaacaa	acaacacaga	ttaaaagcac	360
aatcaccatt	gaaaaaaaaar					380

<210> 773

<211> 472

<212> DNA

<213> Homo sapiens

<400> 773

agatgggtca	ggtgcttcgg	ctgccacggc	agaggttctc	tgctctcttt	tctctgctgg	60
aactgggccc	tagagactgg	ggcaggcggt	tctgactcag	aagatctttt	ggaagaattg	120
ctgtgctggt	tgatgcagtt	gatcactgat	attccactct	tagatattac	atatgaaata	180
tcagtagaag	ctatatcaac	aatggttggt	ttcctttcct	gccaaactct	ccacaaagaa	240
gttttgcgac	agagcatcag	ccacaagtat	ttgatgcgag	gtccatgtct	tcatacacc	300
agcaaaacttg	tgaagacctt	attatataac	tttatcagac	aagaaaagcc	acctcctcca	360
ggggcccatg	ttttccctca	gcagtcggat	gggggaggac	tgctttatgg	acttgcatca	420
ggagtagcaa	caggactctg	gactgtcttc	acactagggtg	gtgtgggcag	ca	472

<210> 774

<211> 521

<212> DNA

<213> Homo sapiens

<400> 774

attccaggcg	cctgcaacta	aacgtggccg	ggtctgcaag	ctaggtgccca	gcggggaaag	60
tttccctgct	tcttatcgtc	tgctttaacg	ccttaaataag	cccgtgaag	gctgcagcag	120
gtgctaggta	gcagcctccc	ggccctcggg	aaaggcgggg	tggggaggcg	agagcagctt	180
agcctcctcg	acctcccttc	ctgggtgacg	acgaacagtt	cccgtagaat	ttcgcttcac	240
cgagtgcact	tgagcccagg	gcgacggtca	gcttggttaat	tcctggctgc	aggaactttg	300
tgagaatttt	aatgcatgga	aaagctgtcc	atgttccaac	tgctgtacat	ccaaaagtct	360
cagtgtgaata	gcaggaccaa	aatattctgt	caatcagctg	accatatact	taatgactcc	420

401

taaaatctcg	tggacttcta	agaaagcgcc	atggcctgtg	ctgctgttat	gattcctggg	480
ttgttgcggt	gctctgttgg	agccatccgt	attgaggctg	y		521

<210> 775
 <211> 447
 <212> DNA
 <213> Homo sapiens

<400> 775						
acttagctga	tagctagcat	tagctgctgg	ccaggtgcat	aagccatctt	agatgtccag	60
cccaggttca	tagaatggga	tgcccactga	tgacaatgcc	aggccctcca	gggaaacgag	120
gtaagagagg	ccgaagagga	gaatctggtc	ctcctggaca	gcctggctct	cagggccctc	180
ctggtccaaa	aggcgataag	ggagaacaag	gtgatcaggg	acctaggatg	stgtttccta	240
aaatcaatca	tgggtttctc	tctgctgatc	agcagctcat	taaacgccgc	ctgattaagg	300
gtgaccaagg	acaggcaggg	cctccaggac	cccctggccc	tccaggccca	agagggccac	360
ctggggacac	agggaaagat	ggccccctgt	gaatgccagg	agtacccggt	gaaccaggaa	420
agccaggaga	acaaggcttg	atgggtc				447

<210> 776
 <211> 491
 <212> DNA
 <213> Homo sapiens

<400> 776						
attctacagg	aagcaatfff	ttggttgtga	agggagaaat	gggctcctca	tggagggtcc	60
tgggctccac	tgaggaagag	cccttcctcc	cagcagtgtt	gtataaagta	gagaagataa	120
atggatggag	aacgcctgct	ccaagccgga	cgacgacatt	ctagacatcc	cgctggacga	180
tcccggcgcc	aacgcggccg	ccgccaaaat	ccaggcgagt	tttcggggcc	acatggcgcg	240
gaagaagata	aagagcggag	actaggccag	aactgagcat	tttcaaagtt	cccgaggaga	300
gatggatgcc	gcgtcccctt	cgcagcgacg	agacttcctt	gcggtgtttg	tgacccccctc	360
ctgcccagca	acctgccagc	tacaggagcc	ccctgcgtcc	cagagactcc	ctcaccacag	420
caggctccgt	cgcgagagtc	ctgagtcctg	gcccttttag	ttagttctgc	agtctagtat	480
ggtcccawtt	g					491

<210> 777
 <211> 389
 <212> DNA
 <213> Homo sapiens

<400> 777						
tgtgtaaatgt	tgaatctgga	aattgatcag	cattaaaggg	cacatgaagc	agtgtctgca	60
ggcgttcagt	gctgcggasc	tgttaaagggt	cactcagatg	tgcagggtgtt	aatcttctct	120
aaaagcctgg	tgatacagct	ctggctttct	gagcacacta	cggatctgga	aaatactgga	180
aaatgtgata	cttagaatac	tttggtgctg	aaggaaactt	cctctccatt	gcagaatagc	240
tgagccaagt	gagtgaagtt	gcagaaagca	ggtggtgagc	tctgcctg	tggagggtgc	300
catggagggc	cattcctgcc	cgccaacagc	accgtcctgc	aggagaccac	ttggcagaag	360
ggtgcagggc	tgctggtgtc	agagcaaga				389

<210> 778
 <211> 440
 <212> DNA
 <213> Homo sapiens

<400> 778						
tgtagtgttt	gcagtagcgt	ttctcacctt	ccctgatgaa	gggtcttttc	agttgcaaac	60
caacttgcca	cctcagcggc	agaagttgag	tgttggtcgt	gtcaccacag	agtctcgttc	120
tgtcaccacg	cctggtgtgc	agtggttcar	gccattctcc	tgctcagcc	tcccaagtag	180
ctgggattac	aggcgccac	caccaagccc	agctaatttt	tgtatgttcg	gtagagacgg	240
ggtttcacca	tgttgccag	gctggtctca	aactcctgac	ctcaagcgac	ccaccgcct	300
cggcctccca	gagttttgcg	attacaggcg	tgacaccacg	cccagcgga	tttcttattc	360
tttttgatct	atcaaaaaca	cctcattgtt	ctcaatgact	gacctataa	taagtaaatg	420
ttatagaaca	taaaaaaaaa					440

<210> 779
 <211> 418
 <212> DNA
 <213> Homo sapiens

<400> 779
 atagattaca acggtgatgg cgggcaagcg gtccggcttg agccgggagg ctctcctcca 60
 gtcctttctc ggcgatgaacc tgggtggtgat gccgcccacc cgggcccagg gtctgcgctt 120
 cggtaccttg gaggggatgc tacagcactg ggaactgggc caggccctgc ggcasgctat 180
 cacggcttcc taaacacctc ttatcaccgg caagagggtt atgtgcgasr gacttgaccg 240
 gactctcatg agtgcgtgagg ccaacctggc tggactcttc cctcccaacg gatgcagcgc 300
 ttcaaccgga acatctcgtg gcagcctatt cctgtgcaca ctgtscatc actgaggaca 360
 ggctgctgaa gttcccggtg ggcccatgtc cccgttatga gcagctgcag aacgagac 418

<210> 780
 <211> 362
 <212> DNA
 <213> Homo sapiens

<400> 780
 aaacttgccc ttctcaaaca tggccgccac ggccgctctg gaagggaacc gctctggggc 60
 ccgcctttga tctcgttggg ggggctgggg gatgagagct gcaccgcgcg ggacaagtcg 120
 ccggcgggcc gacggagcag aagagagagc atggagcttg agaggatcgt cagtgcagcc 180
 ctctttgcct ttgtccagac acacctcccg gagccgacct cagtggcttg gatgagggtca 240
 tcttctccta tgtgcttggg gtcctggagg actggggccc tcggcattca gaggagaact 300
 tcgatatgga ggctttcact gagatgatgg aggcctatgt gcctggcttc gccacatccc 360
 ca 362

<210> 781
 <211> 379
 <212> DNA
 <213> Homo sapiens

<400> 781
 agattatttc tacttggttg ccttccagat gtttcacact tggacagcaa actgatttca 60
 aaccactcyt tttcaaagat ctctgaggga gacattgcac ctggccactg cagcccagag 120
 caggctctggc cacggccatg agcatgctga gccatcatgc ccaccgtgga tgacattctg 180
 gagcaggttg gggagtctgg ctggttccag aagcaagcct tctcatctt atgcctgctg 240
 tcggctgcct ttgcgcccac ctgtgtgggc atcgtcttcc tgggtttcac acctgaccac 300
 cactgccaga gtccctgggt ggctgagctg agccagcgtg gtggctggag ccctgcgagg 360
 agctgaacta tacatgtgcc 379

<210> 782
 <211> 453
 <212> DNA
 <213> Homo sapiens

<400> 782
 ttctcatctt ttgtaatgtg tggtcctcac ttggcctctg ggacacctcc aagtcttagt 60
 tgtcttccta actcactagt tgctccttct cactctcctt tgctgggtct tccttttctc 120
 cttgacctct taatgttaga gtcccagagt gtagtcatag tcttctcttg tgatctatct 180
 ttaccctgtc tttgtgatct agtataatgg ccaccaccta taaacttata gctcacgaat 240
 ttatatcact agctctgaga tctctcacga actccagatg tgcacatctg actgcttact 300
 caacatctcg aatgcctgtt cttcaagagt ttgagtcatg aagtgtgaag cctagagatg 360
 gtgtgatata cagagatgaa actggttttc acaggtgata gatggaaggc tgttttcaaa 420
 agatcaagct tgaccacata ttatcccctc cac 453

<210> 783
 <211> 553
 <212> DNA
 <213> Homo sapiens

<400> 783

403

```

aactacttcc gggggagcgg cgcggcggcg sggagtgtgt tctaaagagt ggtgagtcag      60
aagagacgtc aggcagcaag cgacttgggc catggctctg acctagactt ctcacctccg      120
cagtgcccgga gccacttttc ctggagaacc tgctacggta cggactcttc ctgggagcca      180
tcttccagct catctgtgtg ctggccatca tcgtacccat tcccaagtcc cacgaggcgg      240
aggctgaacc gtctgagccc agaagtgtg aggtgacgag gaagcccaag gctgctgttc      300
cttctgtgaa caagaggccc aagaaagaga ctaagaagaa gcggtagaag aggaggcctg      360
aggagctggg cgggcaggga gaggggtctt gggacagccc tcctgggaat ctacattgtg      420
ttcccccgca ttccaggctc aggggtctgag gaggtgtgta cgccctatga ccgcagagat      480
ctagacagtc gtaacagtcc ccaggctcba gctgggcaat ccaccacttc ctcttccttc      540
tgcttctgtg acg                                     553

```

<210> 784
 <211> 472
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> 445
 <223> n=a, g, c or t

```

<400> 784
ttccctttcg aattccaggg tatatctggg aggccggaga cgtgtctggt tattacacag      60
atgcacagct ggacgtggga tccacacagc tcagaacagt tggatcttgc tcagtctctg      120
tcagaggaag atcccttgga caagaggacs ctgccttggt gtgagagtga ggaagarga      180
agctggaacg agggttaagg aaaaccttcc agtctggaca gtgactggag agctccaagg      240
aaagcccctc ggtaaccacg ccgctggcac catgaacca gagagcagta tctttattga      300
ggattacctt aagtatttcc aggaccaagt gagcagagag aatctgctac aactgctgac      360
tgatgatgaa gcctggaatg gattcggtgg tgctgctgaa ctgccagggg atgaggcaga      420
tgagctccgt aaagctctga acaangcttg caagtcacat ggtcatgaag ga          472

```

<210> 785
 <211> 157
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -29..-1

<220>
 <221> UNSURE
 <222> 120
 <223> Xaa = His,Gln

```

<400> 785
Met Ala Thr Met Val Pro Ser Val Leu Trp Pro Arg Ala Cys Trp Thr
                    -25                    -20                    -15
Leu Leu Val Cys Cys Leu Leu Thr Pro Gly Val Gln Gly Gln Glu Phe
                    -10                    -5                    1
Leu Leu Arg Val Glu Pro Gln Asn Pro Val Leu Ser Ala Gly Gly Ser
    5                    10                    15
Leu Phe Val Asn Cys Ser Thr Asp Cys Pro Ser Ser Glu Lys Ile Ala
20                    25                    30                    35
Leu Glu Thr Ser Leu Ser Lys Glu Leu Val Ala Ser Gly Met Gly Trp
                    40                    45                    50
Ala Ala Phe Asn Leu Ser Asn Val Thr Gly Asn Ser Arg Ile Leu Cys
                    55                    60                    65
Ser Val Tyr Cys Asn Gly Ser Gln Ile Thr Gly Ser Ser Asn Ile Thr
    70                    75                    80
Val Tyr Arg Leu Pro Glu Arg Val Glu Leu Ala Pro Leu Pro Pro Trp
    85                    90                    95
Gln Pro Val Gly Gln Asn Phe Thr Pro Ala Leu Pro Ser Gly Gly Trp

```

404

100 105 110 115
 Val Ala Pro Asp Xaa Pro His Gly Gly Ala Ala Ser Leu
 120 125

<210> 786
 <211> 151
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -16..-1

<400> 786
 Met Val Met Leu Leu Leu Leu Leu Ser Ala Leu Ala Gly Leu Phe Gly
 -15 -10 -5
 Ala Ala Glu Gly Gln Ala Phe His Leu Gly Lys Cys Pro Asn Pro Pro
 1 5 10 15
 Val Gln Glu Asn Phe Asp Val Asn Lys Tyr Leu Gly Arg Trp Tyr Glu
 20 25 30
 Ile Glu Lys Ile Pro Thr Thr Phe Glu Asn Gly Arg Cys Ile Gln Ala
 35 40 45
 Asn Tyr Ser Leu Met Glu Asn Gly Lys Ile Lys Val Leu Asn Gln Glu
 50 55 60
 Leu Arg Ala Asp Gly Thr Val Asn Gln Ile Glu Gly Glu Ala Thr Pro
 65 70 75 80
 Val Asn Leu Thr Glu Pro Ala Lys Leu Glu Val Lys Phe Ser Trp Phe
 85 90 95
 Met Pro Ser Ala Pro Tyr Trp Ile Leu Ala Thr Asp Tyr Glu Asn Tyr
 100 105 110
 Ala Leu Val Tyr Ser Cys Thr Cys Ile Ile Gln Leu Phe His Val Asp
 115 120 125
 Phe Ala Trp Ile Leu Ala Arg
 130 135

<210> 787
 <211> 143
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -17..-1

<400> 787
 Met Ala Leu Glu Val Leu Met Leu Leu Ala Val Leu Ile Trp Thr Gly
 -15 -10 -5
 Ala Glu Asn Leu His Val Lys Ile Ser Cys Ser Leu Asp Trp Leu Met
 1 5 10 15
 Val Ser Val Ile Pro Val Ala Glu Ser Arg Asn Leu Tyr Ile Phe Ala
 20 25 30
 Asp Glu Leu His Leu Gly Met Gly Cys Pro Ala Asn Arg Ile His Thr
 35 40 45
 Tyr Val Tyr Glu Phe Ile Tyr Leu Val Arg Asp Cys Gly Ile Arg Thr
 50 55 60
 Arg Val Val Ser Glu Glu Thr Leu Leu Phe Gln Thr Glu Leu Tyr Phe
 65 70 75
 Thr Pro Arg Asn Ile Asp His Asp Pro Gln Glu Ile His Leu Glu Cys
 80 85 90 95
 Ser Thr Ser Arg Lys Ser Val Trp Leu Thr Pro Val Ser Thr Glu Asn
 100 105 110
 Glu Ile Lys Leu Asp Pro Ser Pro Phe Ile Ala Asp Phe Gln Thr
 115 120 125

<210> 788
 <211> 148
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -17..-1

<400> 788
 Met Ala Leu Glu Val Leu Met Leu Leu Ala Val Leu Ile Trp Thr Gly
 -15 -10 -5
 Ala Glu Asn Leu His Val Lys Ile Ser Cys Ser Leu Asp Trp Leu Met
 1 5 10 15
 Val Ser Val Ile Pro Val Ala Glu Ser Arg Asn Leu Tyr Ile Phe Ala
 20 25 30
 Asp Glu Leu His Leu Gly Met Gly Cys Pro Ala Asn Arg Ile His Thr
 35 40 45
 Tyr Val Tyr Glu Phe Ile Tyr Leu Val Arg Asp Cys Gly Ile Arg Thr
 50 55 60
 Arg Val Val Ser Glu Glu Thr Leu Leu Phe Gln Thr Glu Leu Tyr Phe
 65 70 75
 Thr Pro Arg Asn Ile Asp His Asp Pro Gln Glu Ile His Leu Glu Cys
 80 85 90 95
 Ser Thr Ser Arg Lys Ser Val Trp Leu Thr Pro Val Ser Thr Glu Asn
 100 105 110
 Glu Ile Lys Leu Asp Pro Ser Pro Phe Ile Ala Asp Phe Gln Thr Thr
 115 120 125
 Ala Glu Glu Leu
 130

<210> 789
 <211> 75
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -24..-1

<220>
 <221> UNSURE
 <222> 51
 <223> Xaa = Glu,Lys

<400> 789
 Met Leu Pro Pro Leu Pro Ser Arg Leu Gly Leu Leu Leu Leu Leu Leu
 -20 -15 -10
 Leu Cys Pro Ala His Val Gly Gly Leu Trp Trp Ala Val Gly Ser Pro
 -5 1 5
 Leu Val Met Asp Pro Thr Ser Ile Cys Arg Lys Ala Arg Arg Leu Ala
 10 15 20
 Gly Arg Gln Ala Glu Leu Cys Gln Ala Glu Pro Glu Val Val Ala Glu
 25 30 35 40
 Leu Ala Arg Gly Ala Arg Leu Gly Val Arg Xaa
 45 50

<210> 790
 <211> 64
 <212> PRT
 <213> Homo sapiens

<220>

<221> SIGNAL

<222> -22..-1

<400> 790

Met	Arg	Gly	Met	Lys	Leu	Leu	Gly	Ala	Leu	Leu	Ala	Leu	Ala	Ala	Leu
		-20					-15					-10			
Leu	Gln	Gly	Ala	Val	Ser	Leu	Lys	Ile	Ala	Ala	Phe	Asn	Ile	Gln	Thr
	-5					1				5					10
Phe	Gly	Glu	Thr	Lys	Met	Ser	Asn	Ala	Thr	Leu	Val	Ser	Tyr	Ile	Val
				15					20					25	
Gln	Ile	Leu	Ser	Arg	Tyr	Asp	Ile	Ala	Leu	Val	Gln	Glu	Val	Arg	Asp
			30					35					40		

<210> 791

<211> 117

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -33..-1

<220>

<221> UNSURE

<222> 82

<223> Xaa = Glu,Gly

<400> 791

Met	Ser	Ala	Gln	Glu	Ser	Cys	Leu	Ser	Leu	Ile	Lys	Tyr	Phe	Leu	Phe
			-30					-25					-20		
Val	Phe	Asn	Leu	Phe	Phe	Phe	Val	Leu	Gly	Ser	Leu	Ile	Phe	Cys	Phe
		-15					-10					-5			
Gly	Ile	Trp	Ile	Leu	Ile	Asp	Lys	Thr	Ser	Phe	Val	Ser	Phe	Val	Gly
	1				5					10					15
Leu	Ala	Phe	Val	Pro	Leu	Gln	Ile	Trp	Ser	Lys	Val	Leu	Ala	Ile	Ser
				20					25					30	
Gly	Ile	Phe	Thr	Met	Gly	Ile	Ala	Leu	Gly	Cys	Val	Gly	Ala	Ser	
			35				40					45			
Arg	Ser	Ser	Ala	Ala	Ser	Trp	Ala	Cys	Ile	Leu	Gly	Cys	Cys	Cys	Ser
		50					55					60			
Cys	Leu	Pro	His	Arg	Ser	Pro	Trp	Glu	Ser	Ser	Ser	Pro	Leu	Ser	Gly
	65					70					75				
Pro	Ala	Xaa	Ala	Lys											
															80

<210> 792

<211> 82

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -32..-1

<220>

<221> UNSURE

<222> 35

<223> Xaa = His,Gln

<400> 792

Met	Ala	Lys	Met	Phe	Asp	Leu	Arg	Thr	Lys	Ile	Met	Ile	Gly	Ile	Gly
		-30					-25					-20			

407

Ser Ser Leu Leu Val Ala Ala Met Val Leu Leu Ser Val Val Phe Cys
 -15 -10 -5
 Leu Tyr Phe Lys Val Ala Lys Ala Leu Lys Ala Ala Lys Asp Pro Asp
 1 5 10 15
 Ala Val Ala Val Lys Asn His Asn Pro Asp Lys Val Cys Trp Ala Thr
 20 25 30
 Asn Ser Xaa Ala Lys Ala Thr Thr Met Glu Ser Cys Pro Ser Leu Gln
 35 40 45
 Cys Cys
 50

<210> 793
 <211> 60
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19...-1

<400> 793
 Met Lys Ile Leu Val Ala Leu Ala Val Phe Phe Leu Val Ser Thr Gln
 -15 -10 -5
 Leu Phe Ala Glu Glu Ile Gly Ala Asn Asp Asp Leu Asn Tyr Trp Ser
 1 5 10
 Asp Trp Tyr Asp Ser Asp Gln Ile Lys Glu Glu Leu Pro Glu Pro Phe
 15 20 25
 Glu His Leu Leu Gln Arg Ile Ala Arg Arg Pro Lys
 30 35 40

<210> 794
 <211> 79
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -28...-1

<400> 794
 Met Ile Tyr Thr Met Lys Lys Val His Ala Leu Trp Ala Ser Val Cys
 -25 -20 -15
 Leu Leu Leu Asn Leu Ala Pro Ala Pro Leu Asn Ala Asp Ser Glu Glu
 -10 -5 1
 Asp Glu Glu His Thr Ile Ile Thr Asp Thr Glu Leu Pro Pro Leu Lys
 5 10 15 20
 Leu Met His Ser Phe Cys Ala Phe Lys Ala Asp Asp Gly Pro Cys Lys
 25 30 35
 Ala Ile Met Lys Arg Phe Ser Ser Ile Phe Ser Leu Asp Ser Ala
 40 45 50

<210> 795
 <211> 76
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -24...-1

<400> 795
 Met Gly Ser Ala Cys Ile Lys Val Thr Lys Tyr Phe Leu Phe Leu Phe
 -20 -15 -10

408

Asn Leu Ile Phe Phe Ile Leu Gly Ala Val Ile Leu Gly Phe Gly Val
 -5 1 5
 Trp Ile Leu Ala Asp Lys Ser Ser Phe Ile Ser Val Leu Gln Thr Ser
 10 15 20
 Ser Ser Ser Leu Arg Met Gly Ala Tyr Val Phe Ile Gly Val Gly Ala
 25 30 35 40
 Val Thr Met Leu Met Gly Phe Leu Gly Cys Ile Gly
 45 50

<210> 796
 <211> 151
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -18..-1

<220>
 <221> UNSURE
 <222> 117
 <223> Xaa = Ser,Thr

<400> 796
 Met Lys Ala Leu Ile Val Leu Gly Leu Val Leu Leu Ser Val Thr Val
 -15 -10 -5
 Gln Gly Lys Val Phe Glu Arg Cys Glu Leu Ala Arg Thr Leu Lys Arg
 1 5 10
 Leu Gly Met Asp Gly Tyr Arg Gly Ile Ser Leu Ala Asn Trp Met Cys
 15 20 25 30
 Leu Ala Lys Trp Glu Ser Gly Tyr Asn Thr Arg Ala Thr Asn Tyr Asn
 35 40 45
 Ala Gly Asp Arg Ser Thr Asp Tyr Gly Ile Phe Gln Ile Asn Ser Arg
 50 55 60
 Tyr Trp Cys Asn Asp Gly Lys Thr Pro Gly Ala Val Asn Ala Cys His
 65 70 75
 Leu Ser Cys Ser Ala Leu Leu Gln Asp Asn Ile Ala Asp Ala Val Ala
 80 85 90
 Cys Ala Lys Arg Val Val Arg Asp Pro Pro Thr Ser Ala Ser Gln Ser
 95 100 105 110
 Ala Gly Ile Thr Gly Val Xaa Asn Cys Ala Arg Pro His Ser Val Leu
 115 120 125
 Ile Lys Glu Ile Thr Gln Thr
 130

<210> 797
 <211> 82
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -29..-1

<400> 797
 Met Asp Leu Gln Gly Arg Gly Val Pro Ser Ile Asp Arg Leu Arg Val
 -25 -20 -15
 Leu Leu Met Leu Phe His Thr Met Ala Gln Ile Met Ala Glu Gln Glu
 -10 -5 1
 Val Glu Asn Leu Ser Gly Leu Ser Thr Asn Pro Glu Lys Asp Ile Phe
 5 10 15
 Val Val Arg Glu Asn Gly Thr Thr Cys Leu Met Ala Glu Phe Ala Ala
 20 25 30 35

409

Lys Phe Ile Val Pro Tyr Asp Val Trp Ala Ser Asn Tyr Val Asp Leu
 40 45 50
 Ile Thr

<210> 798
 <211> 60
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -24..-1

<400> 798
 Met Glu Glu Gly Gly Asn Leu Gly Gly Leu Ile Lys Met Val His Leu
 -20 -15 -10
 Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp Val Thr Phe Val
 -5 1 5
 Ser Gly Phe Leu Leu Phe Arg Ser Leu Pro Arg His Thr Phe Gly Leu
 10 15 20
 Val Gln Ser Lys Leu Phe Pro Phe Tyr Phe His Ile
 25 30 35

<210> 799
 <211> 142
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -26..-1

<220>
 <221> UNSURE
 <222> 47
 <223> Xaa = Gly,Arg

<400> 799
 Met Lys Thr Glu Gln Leu Asn Arg Phe Ala Gly Phe Gly Ile Gly Leu
 -25 -20 -15
 Ala Ser Leu Phe Thr Glu Asn Val Leu Ala His Pro Cys Ile Val Leu
 -10 -5 1 5
 Arg Arg Gln Cys Gln Val Asn Tyr His Ala Gln His Tyr His Leu Thr
 10 15 20
 Pro Phe Thr Val Ile Asn Ile Met Tyr Ser Phe Asn Lys Thr Gln Gly
 25 30 35
 Pro Arg Ala Leu Trp Lys Gly Met Xaa Ser Thr Phe Ile Val Gln Gly
 40 45 50
 Val Thr Leu Gly Ala Glu Gly Ile Ile Ser Glu Phe Thr Pro Leu Pro
 55 60 65 70
 Arg Glu Val Leu His Lys Trp Ser Pro Lys Gln Ile Gly Glu His Leu
 75 80 85
 Leu Leu Lys Ser Leu Thr Tyr Val Val Ala Met Leu Phe Tyr Ser Ala
 90 95 100
 Ser Leu Ile Glu Thr Val Gln Ser Glu Ile Ile Arg Asp Asn
 105 110 115

<210> 800
 <211> 126
 <212> PRT
 <213> Homo sapiens

<220>

<221> SIGNAL
 <222> -44...-1

<400> 800

Met	His	Glu	Glu	Glu	Ile	Tyr	Thr	Ser	Leu	Gln	Trp	Asp	Ser	Pro	Ala	
				-40					-35					-30		
Pro	Asp	Thr	Tyr	Gln	Lys	Cys	Leu	Ser	Ser	Asn	Lys	Cys	Ser	Gly	Ala	
			-25					-20					-15			
Cys	Cys	Leu	Val	Met	Val	Ile	Ser	Cys	Val	Phe	Cys	Met	Gly	Leu	Leu	
		-10					-5					1				
Thr	Ala	Ser	Ile	Phe	Leu	Gly	Val	Lys	Leu	Leu	Gln	Val	Ser	Thr	Ile	
5					10					15					20	
Ala	Met	Gln	Gln	Gln	Glu	Lys	Leu	Ile	Gln	Gln	Glu	Arg	Ala	Leu	Leu	
			25						30					35		
Asn	Phe	Thr	Glu	Trp	Lys	Arg	Ser	Cys	Ala	Leu	Gln	Met	Lys	Tyr	Cys	
		40					45						50			
Gln	Ala	Phe	Met	Gln	Asn	Ser	Leu	Ser	Ser	Ala	His	Asn	Ser	Ser	Pro	
		55					60					65				
Cys	Pro	Asn	Asn	Trp	Ile	Gln	Asn	Arg	Glu	Ser	Cys	Tyr	Tyr			
	70					75					80					

<210> 801
 <211> 52
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -20...-1

<400> 801

Met	Gly	Arg	Ala	Arg	Glu	Val	Gly	Trp	Met	Ala	Ala	Gly	Leu	Met	Ile	
-20					-15				-10					-5		
Gly	Ala	Gly	Ala	Cys	Tyr	Cys	Val	Tyr	Lys	Leu	Thr	Ile	Gly	Arg	Asp	
			1					5					10			
Asp	Ser	Glu	Lys	Leu	Glu	Glu	Glu	Gly	Glu	Glu	Glu	Trp	Asp	Asp	Asp	
		15					20					25				
Gln	Glu	Leu	Asp													
		30														

<210> 802
 <211> 154
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -15...-1

<400> 802

Met	Asn	Ile	Leu	Met	Leu	Thr	Phe	Ile	Ile	Cys	Gly	Leu	Leu	Thr	Arg	
-15					-10					-5					1	
Val	Thr	Lys	Gly	Ser	Phe	Glu	Pro	Gln	Lys	Cys	Trp	Lys	Asn	Asn	Val	
			5					10					15			
Gly	His	Cys	Arg	Arg	Arg	Cys	Leu	Asp	Thr	Glu	Arg	Tyr	Ile	Leu	Leu	
		20				25						30				
Cys	Arg	Asn	Lys	Leu	Ser	Cys	Cys	Ile	Ser	Ile	Ile	Ser	His	Glu	Tyr	
		35				40					45					
Thr	Arg	Arg	Pro	Ala	Phe	Pro	Val	Ile	His	Leu	Glu	Asp	Ile	Thr	Leu	
50					55					60					65	
Asp	Tyr	Ser	Asp	Val	Asp	Ser	Phe	Thr	Gly	Ser	Pro	Val	Ser	Met	Leu	
			70						75					80		
Asn	Asp	Leu	Ile	Thr	Phe	Asp	Thr	Thr	Lys	Phe	Gly	Glu	Thr	Met	Thr	

411

			85					90					95				
Pro	Glu	Thr	Asn	Thr	Pro	Glu	Thr	Met	Pro	Pro	Ser	Glu	Ala	Thr			
			100				105				110						
Thr	Pro	Glu	Thr	Thr	Met	Pro	Pro	Ser	Glu	Thr	Ala	Thr	Ser	Glu	Thr		
			115				120				125						
Met	Pro	Pro	Pro	Ser	Gln	Thr	Ala	Leu	Thr								
130						135											

<210> 803

<211> 165

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -39..-1

<400> 803

Met	Met	Thr	Ile	Thr	Phe	Leu	Pro	Tyr	Thr	Phe	Ser	Leu	Met	Val	Thr		
				-35					-30					-25			
Phe	Pro	Asp	Val	Pro	Leu	Gly	Ile	Phe	Leu	Phe	Cys	Val	Cys	Val	Ile		
			-20					-15					-10				
Ala	Ile	Gly	Val	Val	Gln	Ala	Leu	Ile	Val	Gly	Tyr	Ala	Phe	His	Phe		
			-5				1				5						
Pro	His	Leu	Leu	Ser	Pro	Gln	Ile	Gln	Arg	Ser	Ala	His	Arg	Ala	Leu		
10					15					20				25			
Tyr	Arg	Arg	His	Val	Leu	Gly	Ile	Val	Leu	Gln	Gly	Pro	Ala	Leu	Cys		
				30					35					40			
Phe	Ala	Ala	Ala	Ile	Phe	Ser	Leu	Phe	Phe	Val	Pro	Leu	Ser	Tyr	Leu		
			45					50					55				
Leu	Met	Val	Thr	Val	Ile	Leu	Leu	Pro	Tyr	Val	Ser	Lys	Val	Thr	Gly		
			60				65					70					
Trp	Cys	Arg	Asp	Arg	Leu	Leu	Gly	His	Arg	Glu	Pro	Ser	Ala	His	Pro		
75					80					85							
Val	Glu	Val	Phe	Ser	Phe	Asp	Leu	His	Glu	Pro	Leu	Ser	Lys	Glu	Arg		
90					95					100					105		
Val	Glu	Ala	Phe	Ser	Asp	Gly	Val	Tyr	Ala	Ile	Val	Ala	Thr	Leu	Leu		
				110					115					120			
Ile	Leu	Asp	Ile	Trp													
				125													

<210> 804

<211> 145

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -22..-1

<220>

<221> UNSURE

<222> 84

<223> Xaa = Ala,Pro,Ser,Thr

<220>

<221> UNSURE

<222> -1

<223> Xaa = Ala,Ser

<220>

<221> UNSURE

<222> 21

<223> Xaa = Cys, Phe, Ile, Asn, Ser, Thr, Tyr

<220>

<221> UNSURE

<222> 46

<223> Xaa = Cys, Phe, Ser, Tyr

<400> 804

Met	Gly	Arg	Trp	Ser	Leu	Thr	Ala	Ser	Pro	Val	Thr	Leu	Thr	Ser	Leu
		-20					-15					-10			
Leu	Pro	Met	Val	Thr	Xaa	Gly	Pro	Ala	Ala	Gly	Lys	Ser	Ser	Ser	
	-5					1				5					10
Thr	Ser	Trp	Asp	Thr	Val	Leu	Thr	Ala	Ile	Xaa	Cys	Ala	Ser	Thr	Val
			15					20						25	
Gln	Leu	Ile	Ser	Thr	Thr	Leu	Gly	Ala	Ser	Ala	Ser	Gly	Ala	Arg	Met
			30					35					40		
Pro	Thr	Thr	Xaa	Cys	Ser	Gly	Thr	Thr	Val	Phe	Leu	Thr	Ala	Leu	Gln
	45						50					55			
Asp	Thr	Met	Gln	Arg	Glu	Glu	Leu	Val	Lys	Asn	Ala	Thr	Pro	Pro	Ala
	60					65				70					
Glu	Pro	Ala	Arg	Ala	Glu	Asp	Leu	Ser	Xaa	Ala	Pro	His	Val	Thr	Pro
	75				80					85					90
Thr	Ser	Cys	Cys	Pro	Thr	Leu	Ala	Pro	Ala	Ala	Pro	Pro	Ala	Ser	Leu
				95					100					105	
Gly	Thr	Ile	Leu	Met	Thr	Ile	Met	Phe	Ala	Ser	Val	Gly	Phe	Ser	Asn
			110					115					120		

Glu

<210> 805

<211> 109

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -30..-1

<220>

<221> UNSURE

<222> 64

<223> Xaa = Ile, Met

<220>

<221> UNSURE

<222> 58

<223> Xaa = Leu, Val

<220>

<221> UNSURE

<222> 42

<223> Xaa = Lys, Arg

<400> 805

Met	Ala	Gly	Lys	Arg	Ser	Gly	Trp	Ser	Arg	Ala	Ala	Leu	Leu	Gln	Leu
	-30				-25					-20					-15
Leu	Leu	Gly	Val	Asn	Leu	Val	Val	Met	Pro	Pro	Thr	Arg	Ala	Arg	Ser
			-10						-5					1	
Leu	Arg	Phe	Val	Thr	Leu	Glu	Gly	Met	Leu	Gln	His	Trp	Glu	Leu	Gly
	5						10				15				
Gln	Ala	Leu	Arg	Gln	Arg	Tyr	His	Gly	Phe	Leu	Asn	Thr	Ser	Tyr	His
	20					25					30				
Arg	Gln	Glu	Val	Tyr	Val	Arg	Xaa	Tyr	Phe	Asp	Arg	Thr	Leu	Met	Ser
	35				40					45					50

413

Ala Glu Ala Asn Leu Ala Gly Xaa Phe Pro Pro Asn Gly Xaa Gln Arg
55 60 65
Phe Asn Pro Asn Ile Ser Trp Gln Pro Ile Pro Val His
70 75

```
<210> 806
<211> 57
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -32..-1
```

```

<400> 806
Met Ala Thr Pro Pro Phe Arg Leu Ile Arg Lys Met Phe Ser Phe Lys
      -30                -25                -20
Val Ser Arg Trp Met Gly Leu Ala Cys Phe Arg Ser Leu Ala Ala Ser
      -15                -10                -5
Ser Pro Ser Ile Arg Gln Lys Lys Leu Met His Lys Leu Gln Glu Glu
1      5      10      15
Asn Ala Phe Arg Glu Glu Met Lys Ile
      20                25

```

```
<210> 807
<211> 79
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -19..-1
```

<400> 807															
Met	Val	Gly	Ile	Leu	Pro	Leu	Cys	Cys	Ser	Gly	Cys	Val	Pro	Ser	Leu
				-15					-10					-5	
Cys	Cys	Ser	Ser	Tyr	Val	Pro	Ser	Val	Ala	Pro	Thr	Ala	Ala	His	Ser
			1				5					10			
Val	Arg	Val	Pro	His	Ser	Ala	Gly	His	Cys	Gly	Gln	Arg	Val	Leu	Ala
	15					20					25				
Cys	Ser	Leu	Pro	Gln	Val	Phe	Leu	Lys	Pro	Trp	Ile	Phe	Val	Glu	His
	30				35					40					45
Phe	Ser	Ser	Trp	Leu	Ser	Leu	Glu	Leu	Phe	Ser	Phe	Leu	Arg	Tyr	
				50					55					60	

```
<210> 808
<211> 147
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -20..-1
```

<400> 808															
Met	Asp	Ser	Ser	Thr	Ala	His	Ser	Pro	Val	Phe	Leu	Val	Phe	Pro	Pro
					-15					-10					-5
Glu	Ile	Thr	Ala	Ser	Glu	Tyr	Glu	Ser	Thr	Glu	Leu	Ser	Ala	Thr	Thr
				1					5					10	
Phe	Ser	Thr	Gln	Ser	Pro	Leu	Gln	Lys	Leu	Phe	Ala	Arg	Lys	Met	Lys
				15					20					25	
Ile	Leu	Gly	Thr	Ile	Gln	Ile	Leu	Phe	Gly	Ile	Met	Thr	Phe	Ser	Phe
				30					35					40	

414

Gly Val Ile Phe Leu Phe Thr Leu Leu Lys Pro Tyr Pro Arg Phe Pro
 45 50 55 60
 Phe Ile Phe Leu Ser Gly Tyr Pro Phe Trp Gly Ser Val Leu Phe Ile
 65 70 75
 Asn Ser Gly Ala Phe Leu Ile Ala Val Lys Arg Lys Thr Thr Glu Thr
 80 85 90
 Leu Ile Ile Leu Ser Arg Ile Met Asn Phe Leu Ser Ala Leu Gly Ala
 95 100 105
 Ile Ala Gly Ile Ile Leu Leu Thr Phe Gly Phe His Pro Arg Ser Lys
 110 115 120
 Leu His Leu
 125

<210> 809
 <211> 96
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -33..-1

<400> 809
 Met Lys Leu Lys Leu Lys Asn Val Phe Leu Ala Tyr Phe Leu Val Ser
 -30 -25 -20
 Ile Ala Gly Leu Leu Tyr Ala Leu Val Gln Leu Gly Gln Pro Cys Asp
 -15 -10 -5
 Cys Leu Pro Pro Leu Arg Ala Ala Glu Gln Leu Arg Gln Lys Asp
 1 5 10 15
 Leu Arg Ile Ser Gln Leu Gln Ala Glu Leu Arg Arg Pro Pro Pro Ala
 20 25 30
 Pro Ala Gln Pro Pro Glu Pro Glu Ala Leu Pro Thr Ile Tyr Val Val
 35 40 45
 Thr Pro Thr Tyr Ala Arg Leu Val Gln Lys Ala Glu Leu Val Arg Leu
 50 55 60

<210> 810
 <211> 169
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -16..-1

<400> 810
 Met Ser Gly Cys Gly Leu Phe Leu Arg Thr Thr Ala Ala Ala Arg Ala
 -15 -10 -5
 Cys Arg Gly Leu Val Val Ser Thr Ala Asn Arg Arg Leu Leu Arg Thr
 1 5 10 15
 Ser Pro Pro Val Arg Ala Phe Ala Lys Glu Leu Phe Leu Gly Lys Ile
 20 25 30
 Lys Lys Lys Glu Val Phe Pro Phe Pro Glu Val Ser Gln Asp Glu Leu
 35 40 45
 Asn Glu Ile Asn Gln Phe Leu Gly Pro Val Glu Lys Phe Phe Thr Glu
 50 55 60
 Glu Val Asp Ser Arg Lys Ile Asp Gln Glu Gly Lys Ile Pro Asp Glu
 65 70 75 80
 Thr Leu Glu Lys Leu Lys Ser Leu Gly Leu Phe Gly Leu Gln Val Pro
 85 90 95
 Glu Glu Tyr Gly Gly Leu Gly Phe Ser Asn Thr Met Tyr Ser Arg Leu
 100 105 110
 Gly Glu Ile Ile Ser Met Asp Gly Ser Ile Thr Val Thr Leu Ala Ala

415

[illegible]

```
<210> 811
<211> 136
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SIGNAL
<222> -36..-1
```

<400>	811															
Met	Glu	Glu	Leu	Gln	Glu	Pro	Leu	Arg	Gly	Glu	Leu	Arg	Leu	Cys	Phe	
	-35					-30					-25					
Thr	Gln	Ala	Ala	Arg	Thr	Ser	Leu	Leu	Leu	Leu	Arg	Leu	Asn	Asp	Ala	
-20					-15					-10					-5	
Ala	Leu	Arg	Ala	Leu	Gln	Glu	Cys	Gln	Arg	Gln	Gln	Val	Arg	Pro	Val	
				1			5						10			
Ile	Ala	Phe	Gln	Gly	His	Arg	Gly	Tyr	Leu	Arg	Leu	Pro	Gly	Pro	Gly	
	15						20					25				
Trp	Ser	Cys	Leu	Phe	Ser	Phe	Ile	Val	Ser	Gln	Cys	Cys	Gln	Glu	Gly	
	30					35					40					
Ala	Gly	Gly	Ser	Leu	Asp	Leu	Val	Cys	Gln	Arg	Phe	Leu	Arg	Ser	Gly	
45					50					55					60	
Pro	Asn	Ser	Leu	His	Cys	Leu	Gly	Ser	Leu	Arg	Glu	Arg	Leu	Ile	Ile	
				65					70					75		
Trp	Ala	Ala	Met	Asp	Ser	Ile	Pro	Ala	Pro	Ser	Ser	Val	Gln	Gly	His	
			80					85					90			
Asn	Leu	Thr	Glu	Asp	Ala	Arg	His									
		95					100									

```
<210> 812
<211> 109
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> SIGNAL  
<222> -23..-1
```

```
<220>
<221> UNSURE
<222> 81
<223> Xaa = Pro,Ser
```

<400> 812															
Met	Cys	Tyr	Gly	Lys	Cys	Ala	Arg	Cys	Ile	Gly	His	Ser	Leu	Val	Gly
			-20					-15					-10		
Leu	Ala	Leu	Leu	Cys	Ile	Ala	Ala	Asn	Ile	Leu	Leu	Tyr	Phe	Pro	Asn
		-5					1				5				
Gly	Glu	Thr	Lys	Tyr	Ala	Ser	Glu	Asn	His	Leu	Ser	Arg	Phe	Val	Trp
10					15					20					25
Phe	Phe	Ser	Gly	Ile	Val	Gly	Gly	Gly	Leu	Leu	Met	Leu	Leu	Pro	Ala
				30					35					40	
Phe	Val	Phe	Ile	Gly	Leu	Glu	Gln	Asp	Asp	Cys	Cys	Gly	Cys	Cys	Gly
			45					50					55		
His	Glu	Asn	Cys	Gly	Lys	Arg	Cys	Ala	Met	Leu	Ser	Ser	Val	Leu	Ala
		60					65					70			
Ala	Leu	Ile	Gly	Ile	Ala	Gly	Xaa	Gly	Tyr	Cys	Val	Ile			

75

80

85

<210> 813
 <211> 74
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -17..-1

<400> 813
 Met Trp Pro Asn Leu Leu Pro Leu Cys Gln Arg Ile His Cys Ile Ile
 -15 -10 -5
 Gly Arg Phe Arg Cys Asn Gly Phe Glu Asp Cys Pro Asp Gly Ser Asp
 1 5 10 15
 Glu Glu Asn Cys Thr Ala Asn Pro Leu Leu Cys Ser Thr Ala Arg Tyr
 20 25 30
 His Cys Lys Asn Gly Leu Cys Ile Asp Lys Ser Phe Ile Cys Asp Gly
 35 40 45
 Gln Asn Asn Cys Gln Asp Asn Ser Asp Glu
 50 55

<210> 814
 <211> 137
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

<220>
 <221> UNSURE
 <222> 87
 <223> Xaa = Pro,Arg

<400> 814
 Met Thr Leu Gln Trp His Leu Tyr Glu Met Ala Arg Asn Leu Lys Val
 -45 -40 -35
 Gln Asp Met Leu Arg Ala Glu Val Leu Ala Ala Arg His Gln Ala Gln
 -30 -25 -20
 Gly Asp Met Ala Thr Met Leu Gln Leu Val Pro Leu Leu Lys Ala Ser
 -15 -10 -5 1
 Ile Lys Glu Thr Leu Arg Leu His Pro Ile Ser Val Thr Leu Gln Arg
 5 10 15
 Tyr Leu Val Asn Asp Leu Val Leu Arg Asp Tyr Met Ile Pro Ala Lys
 20 25 30
 Thr Leu Val Gln Val Ala Ile Tyr Ala Leu Gly Arg Glu Pro Thr Phe
 35 40 45
 Phe Phe Asp Pro Glu Asn Phe Asp Pro Thr Arg Trp Leu Ser Lys Asp
 50 55 60 65
 Lys Asn Ile Thr Tyr Phe Arg Asn Leu Gly Phe Gly Trp Gly Val Arg
 70 75 80
 Gln Cys Leu Gly Arg Xaa Ile Ala Glu
 85 90

<210> 815
 <211> 98
 <212> PRT
 <213> Homo sapiens

<220>

<221> SIGNAL
 <222> -20..-1

<400> 815

Met	Asp	Lys	Asp	Ser	Gln	Gly	Leu	Leu	Asp	Ser	Ser	Leu	Met	Ala	Ser
-20					-15				-10						-5
Gly	Thr	Ala	Ser	Arg	Ser	Glu	Asp	Glu	Glu	Ser	Leu	Ala	Gly	Gln	Lys
			1				5						10		
Arg	Ala	Ser	Ser	Gln	Ala	Leu	Gly	Thr	Ile	Pro	Lys	Arg	Arg	Ser	Ser
	15					20					25				
Ser	Arg	Phe	Ile	Lys	Arg	Lys	Lys	Phe	Asp	Asp	Glu	Leu	Val	Glu	Ser
30						35				40					
Ser	Leu	Ala	Lys	Ser	Ser	Thr	Arg	Ala	Lys	Gly	Ala	Ser	Gly	Val	Glu
45					50					55					60
Gln	Gly	Ala	Val	Arg	Val	Val	Asn	Pro	Pro	Pro	Val	Arg	Arg	Arg	Arg
				65					70					75	

Tyr Gln

<210> 816
 <211> 113
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -36..-1

<400> 816

Met	Met	Thr	Glu	Lys	His	Glu	Asp	His	Cys	Leu	Lys	Ser	Cys	Ser	Gly
-35						-30					-25				
His	Ser	Tyr	Ile	Arg	Lys	Asn	Trp	Leu	Gly	Cys	Ile	Val	Phe	Pro	Phe
-20					-15				-10						-5
Ser	Ala	Leu	Leu	Gln	Ser	Glu	Ile	Ser	Gly	Thr	Phe	Gln	Val	Thr	
			1				5					10			
Ile	Pro	Pro	Val	Leu	Leu	Gly	Tyr	Thr	Trp	Ser	Asn	Thr	Tyr	Val	Phe
	15					20					25				
Pro	Lys	Glu	Asp	Ser	Asn	Glu	Gln	Asn	Leu	Lys	Glu	Cys	Thr	Phe	Leu
30						35				40					
Asn	Ile	Phe	Ala	Thr	Ile	Glu	Pro	Gln	Ile	Ser	Tyr	Val	Thr	Cys	Asn
45					50					55					60
Pro	Thr	Leu	Asp	Lys	Phe	Leu	Asp	Gln	Thr	Glu	Val	Leu	Gln	Arg	Ala
				65					70					75	

Gln

<210> 817
 <211> 136
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -41..-1

<220>
 <221> UNSURE
 <222> 91
 <223> Xaa = Phe,Leu

<400> 817

Met	Lys	Ala	Ile	Lys	Lys	Ser	Leu	Thr	Glu	Glu	Glu	Tyr	Leu	Tyr	Leu
-40						-35					-30				
Asp	Phe	Ser	His	Gln	Thr	Glu	Gly	Cys	Ile	Phe	Pro	Leu	His	Thr	Ser
-25					-20					-15					-10

418

```

Val Thr Leu Phe Leu Leu Ser Tyr Cys Asp Cys Lys Ile Phe Lys Ile
              -5              1              5
Cys Leu Val Val Thr Lys Glu Val Ser Arg Asp Ser Ser Leu Leu Arg
      10              15              20
Asp Asp Leu Ile Gln Asp Val Glu Ile Gln Ile Ile Ser Arg Gln Glu
      25              30              35
Leu Pro Pro Ile Val Gln Asn Cys Cys Leu Pro Ala Val Val Glu Arg
40              45              50              55
Ser Asp Asn Phe Cys Arg Ala Gly Leu Ala Val Val Leu Arg His Ile
              60              65              70
Ile Gln Lys Ser Tyr Glu Ala Asp Pro Leu Lys Lys Glu Leu Leu Glu
              75              80              85
Leu Leu Gly Xaa Lys Arg Leu Ala
      90              95

```

<210> 818

<211> 137

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -16..-1

<220>

<221> UNSURE

<222> 57

<223> Xaa = Asp,Gly

<220>

<221> UNSURE

<222> 58

<223> Xaa = Cys,Ser

<400> 818

```

Met Glu Asn Phe Ser Leu Leu Ser Ile Ser Gly Pro Pro Ile Ser Ser
      -15              -10              -5
Ser Ala Leu Ser Ala Phe Pro Asp Ile Met Phe Ser Arg Ala Thr Ser
1              5              10              15
Leu Pro Asp Ile Ala Lys Thr Ala Val Pro Thr Glu Ala Ser Ser Pro
      20              25              30
Ala Gln Ala Leu Pro Pro Gln Tyr Gln Ser Ile Ile Val Arg Gln Gly
      35              40              45
Ile Gln Asn Thr Val Leu Ser Pro Xaa Xaa Ser Leu Gly Asp Thr Gln
      50              55              60
His Gly Glu Lys Leu Arg Arg Asn Cys Thr Ile Tyr Arg Pro Trp Phe
65              70              75              80
Ser Pro Tyr Ser Tyr Phe Val Cys Ala Asp Lys Glu Ser Gln Leu Glu
              85              90              95
Ala Tyr Asp Phe Pro Glu Val Gln Gln Asp Glu Gly Lys Trp Asp Asn
      100              105              110
Cys Leu Ser Glu Asp Met Ala Glu Asn
      115              120

```

<210> 819

<211> 113

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 34

<223> Xaa = Ala,Glu

<400> 819

```

Met Leu Gly Ser Arg Val Asp Arg Pro Lys Leu Ser Val Ala Pro Ser
1      5      10      15
Val Val Leu Glu Glu Asp Gln Val Leu Val Ser Pro Ala Val Asp Leu
20      25      30
Glu Xaa Gly Cys Arg Leu Arg Asp Phe Thr Glu Lys Ile Met Asn Val
35      40      45
Lys Gly Lys Val Ile Leu Ser Met Leu Val Val Ser Thr Val Ile Ile
50      55      60
Val Phe Trp Glu Phe Ile Asn Ser Thr Glu Gly Ser Phe Leu Trp Ile
65      70      75      80
Tyr His Ser Lys Asn Pro Glu Val Asp Asp Ser Ser Ala Gln Lys Gly
85      90      95
Trp Trp Phe Leu Ser Trp Phe Asn Asn Gly Ile His Asn Tyr Gln Gln
100      105      110
Gly

```

<210> 820

<211> 174

<212> PRT

<213> Homo sapiens

<400> 820

```

Met Ala Ala Gly Leu Phe Gly Leu Ser Ala Arg Arg Leu Leu Ala Ala
1      5      10      15
Ala Ala Thr Arg Gly Leu Pro Ala Ala Arg Val Arg Trp Glu Ser Ser
20      25      30
Phe Ser Arg Thr Val Val Ala Pro Ser Ala Val Ala Gly Lys Arg Pro
35      40      45
Pro Glu Pro Thr Thr Pro Trp Gln Glu Asp Pro Glu Pro Glu Asp Glu
50      55      60
Asn Leu Tyr Glu Lys Asn Pro Asp Ser His Gly Tyr Asp Lys Asp Pro
65      70      75      80
Val Leu Asp Val Trp Asn Met Arg Leu Val Phe Phe Phe Gly Val Ser
85      90      95
Ile Ile Leu Val Leu Gly Ser Thr Phe Val Ala Tyr Leu Pro Asp Tyr
100      105      110
Arg Met Lys Glu Trp Ser Arg Arg Glu Ala Glu Arg Leu Val Lys Tyr
115      120      125
Arg Glu Ala Asn Gly Phe Pro Ser Trp Asn Pro Thr Ala Ser Thr Pro
130      135      140
Ala Arg Ser Ser Cys Gln Arg Met Ser Asp Gln Leu Leu Ser Gly Ala
145      150      155      160
Gln Glu Ala Pro Pro Ser Pro Pro Pro Ala Cys His Ser Asp
165      170

```

<210> 821

<211> 159

<212> PRT

<213> Homo sapiens

<400> 821

```

Met Glu Arg Tyr Ala Ala Ala Leu Glu Glu Val Ala Asp Gly Ala Arg
1      5      10      15
Gln Gln Glu Arg His Tyr Gln Leu Leu Ser Ala Leu Gln Ser Leu Val
20      25      30
Lys Glu Leu Pro Ser Ser Phe Gln Gln Arg Leu Ser Tyr Thr Thr Leu
35      40      45
Ser Asp Leu Ala Leu Ala Leu Leu Asp Gly Thr Val Phe Glu Ile Val
50      55      60
Gln Gly Leu Leu Glu Ile Gln His Leu Thr Glu Lys Ser Leu Tyr Asn
65      70      75      80

```

420

Gln Arg Leu Arg Leu Gln Asn Glu His Arg Val Leu Arg Gln Ala Leu
 85 90 95
 Arg Gln Lys His Gln Glu Ala Gln Gln Ala Cys Arg Pro His Asn Leu
 100 105 110
 Pro Val Val Gln Ala Ala Gln Gln Arg Glu Leu Glu Ala Val Glu His
 115 120 125
 Arg Ile Arg Glu Glu Gln Arg Ala Met Asp His Lys Ile Ile Leu Glu
 130 135 140
 Leu Asp Arg Lys Val Ala Asp Gln Gln Ser Thr Leu Glu Lys Ala
 145 150 155

<210> 822

<211> 143

<212> PRT

<213> Homo sapiens

<400> 822

Met Ser Gly Phe Glu Asn Leu Asn Thr Asp Phe Tyr Gln Thr Ser Tyr
 1 5 10 15
 Ser Ile Asp Asp Gln Ser Gln Gln Ser Tyr Asp Tyr Gly Gly Ser Gly
 20 25 30
 Gly Pro Tyr Ser Lys Gln Tyr Ala Gly Tyr Asp Tyr Ser Gln Gln Gly
 35 40 45
 Arg Phe Val Pro Pro Asp Met Met Gln Pro Gln Gln Pro Tyr Thr Gly
 50 55 60
 Gln Ile Tyr Gln Pro Thr Gln Ala Tyr Thr Pro Ala Ser Pro Gln Pro
 65 70 75 80
 Phe Tyr Gly Asn Asn Phe Glu Asp Glu Pro Pro Leu Leu Glu Glu Leu
 85 90 95
 Gly Ile Asn Phe Asp His Ile Trp Gln Lys Thr Leu Thr Val Leu His
 100 105 110
 Pro Leu Lys Val Ala Asp Gly Ser Ile Met Asn Glu Thr Asp Leu Ala
 115 120 125
 Gly Pro Met Val Phe Cys Leu Ala Phe Trp Ser Thr Leu Leu Leu
 130 135 140

<210> 823

<211> 138

<212> PRT

<213> Homo sapiens

<400> 823

Met Val Lys Thr Lys Leu Lys Thr Val Val Ile Asp Ile Thr Val Ile
 1 5 10 15
 Pro Phe Ala Asp Asp Phe Phe Cys Pro Gly Ala Pro Ala Gly Val Glu
 20 25 30
 Ile Ser Pro Ala Ala Asp Gly Pro Ser Pro Glu Gly Pro Ile Ala Glu
 35 40 45
 Pro Asp Gly Glu Ser Asp Gly Glu Glu Glu Val Glu Pro Pro Gly Ala
 50 55 60
 Asn Asp Gly Glu Thr Val Gly Glu Phe Glu Thr Gly Asp Glu Asp Leu
 65 70 75 80
 Gly Ala Ile Asp Pro Gly Gly Glu Val Ile Gly Ala Glu Asp Gly Gly
 85 90 95
 Ile Val Glu Asp Asp Leu Gly Ala Ile Gly Ala Gly Gly Glu Val Ile
 100 105 110
 Gly Ala Glu Asp Gly Glu Thr Gly Asp Val Asp Leu Gly Gly Ala Gly
 115 120 125
 Gly Glu Val Val Gly Glu Thr Gly Glu Leu
 130 135

<210> 824

<211> 74

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 53

<223> Xaa = * ,Ala,Asp,Glu,Phe,Leu,Ser,Val,Tyr

<220>

<221> UNSURE

<222> 28

<223> Xaa = His,Asn,Tyr

<220>

<221> UNSURE

<222> 26

<223> Xaa = Ile,Lys,Met,Asn

<220>

<221> UNSURE

<222> 27

<223> Xaa = any one of the twenty amino acids

<400> 824

Met	Arg	Ser	His	Thr	Ile	Thr	Met	Thr	Thr	Thr	Ser	Val	Ser	Ser	Trp
1			5					10					15		
Pro	Tyr	Ser	Ser	His	Arg	Met	Arg	Phe	Xaa	Xaa	Xaa	His	Ser	Asp	Gln
			20					25				30			
Pro	Pro	Gln	Asn	Phe	Ser	Ala	Thr	Pro	Asn	Val	Thr	Thr	Cys	Pro	Met
		35				40					45				
Asp	Glu	Lys	Leu	Xaa	Thr	Thr	Val	Leu	Thr	Thr	Ser	Tyr	Ser	Val	Ile
	50					55					60				
Phe	Ile	Val	Gly	Leu	Val	Gly	Asn	Ile	Ile						
65					70										

<210> 825

<211> 135

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 8

<223> Xaa = Glu,Gln

<220>

<221> UNSURE

<222> 63

<223> Xaa = Glu,Gly

<220>

<221> UNSURE

<222> 133

<223> Xaa = His,Pro

<400> 825

Met	Lys	Glu	Ala	Gly	Gln	Met	Xaa	Asn	Leu	Glu	Ser	Ala	Arg	Ala	Gly
1			5					10					15		
Arg	Ser	Val	Ser	Thr	Gln	Thr	Gly	Ser	Met	Thr	Gly	Gln	Ile	Pro	Arg
			20					25				30			
Leu	Ser	Lys	Val	Asn	Leu	Phe	Thr	Leu	Leu	Ser	Leu	Trp	Met	Glu	Leu
		35				40					45				
Phe	Pro	Ala	Glu	Ala	Gln	Arg	Gln	Lys	Ser	Gln	Lys	Asn	Glu	Xaa	Gly

422

50 55 60
 Lys His Gly Pro Leu Gly Asp Asn Glu Glu Arg Thr Arg Val Ser Thr
 65 70 75 80
 Asp Lys Arg Gln Val Lys Arg Thr Gly Leu Val Val Val Lys Asn Met
 85 90 95
 Lys Ile Val Gly Leu His Cys Ser Ser Glu Asp Leu His Ala Gly Gln
 100 105 110
 Ile Ala Leu Ile Lys His Gly Ser Arg Leu Lys Asn Cys Asp Pro Tyr
 115 120 125
 Phe Ser Arg Lys Xaa Val Leu
 130 135

<210> 826

<211> 61

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 44

<223> Xaa = * ,Glu,Gln

<400> 826

Met Pro Gly Ala Ala Gly Arg Arg Cys Pro Asp Arg Ser Pro Pro Phe
 1 5 10 15
 Gly Pro Asp Phe Gly Ser Ser Gly Val Leu Met Leu Leu Lys Pro Glu
 20 25 30
 Ser Thr Cys Pro Asp Pro Ser Leu Tyr Asp Ala Xaa Leu Cys Pro Val
 35 40 45
 Pro Glu Pro Lys Pro Cys Arg Gly Val Ala Ser Asp Pro
 50 55 60

<210> 827

<211> 130

<212> PRT

<213> Homo sapiens

<400> 827

Met Arg Ser His Thr Ile Thr Met Thr Thr Thr Ser Val Ser Ser Trp
 1 5 10 15
 Pro Tyr Ser Ser His Arg Met Arg Phe Ile Thr Asn His Ser Asp Gln
 20 25 30
 Pro Pro Gln Asn Phe Ser Ala Thr Pro Asn Val Thr Thr Cys Pro Met
 35 40 45
 Asp Glu Lys Leu Leu Ser Thr Val Leu Thr Thr Ser Tyr Ser Val Ile
 50 55 60
 Phe Ile Val Gly Leu Val Gly Asn Ile Ile Ala Leu Tyr Val Phe Leu
 65 70 75 80
 Gly Ile His Arg Lys Arg Asn Ser Ile Gln Ile Tyr Leu Leu Asn Val
 85 90 95
 Ala Ile Ala Asp Leu Leu Leu Ile Phe Cys Leu Pro Phe Arg Ile Met
 100 105 110
 Tyr His Ile Asn Gln Asn Lys Trp Thr Leu Gly Val Ile Leu Cys Lys
 115 120 125
 Val Val
 130

<210> 828

<211> 67

<212> PRT

<213> Homo sapiens

<400> 828

423

Met Asn Phe Leu Ser Ala Leu Gly Ala Ile Ala Gly Ile Ile Leu Leu
 1 5 10 15
 Thr Phe Gly Phe Ile Leu Asp Gln Asn Tyr Ile Cys Gly Tyr Ser His
 20 25 30
 Gln Asn Ser Gln Cys Lys Ala Val Thr Val Leu Phe Leu Gly Ile Leu
 35 40 45
 Ile Thr Leu Met Thr Phe Ser Ile Ile Glu Leu Phe Ile Ser Leu Leu
 50 55 60
 Ser Gln Phe
 65

<210> 829
 <211> 55
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 26
 <223> Xaa = His,Gln

<400> 829
 Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg Gln
 1 5 10 15
 Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser Arg
 20 25 30
 Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala Gly
 35 40 45
 Asp Pro Lys Lys Lys Lys Lys
 50 55

<210> 830
 <211> 55
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 26
 <223> Xaa = His,Gln

<400> 830
 Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg Gln
 1 5 10 15
 Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser Arg
 20 25 30
 Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala Gly
 35 40 45
 Asp Pro Lys Lys Lys Lys Lys
 50 55

<210> 831
 <211> 86
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 9,72
 <223> Xaa = Gly,Arg,Trp

<220>
 <221> UNSURE

<222> 67
 <223> Xaa = Lys, Met, Arg, Thr

<220>
 <221> UNSURE
 <222> 2
 <223> Xaa = Phe, Leu

<220>
 <221> UNSURE
 <222> 85
 <223> Xaa = Ser, Tyr

<400> 831
 Met Xaa Leu Val Asp Trp Phe Tyr Xaa Val Leu Ser Ser Leu Gly Leu
 1 5 10 15
 Trp Gln Lys Glu Ala Lys Ile Leu Phe Leu Gly Leu Asp Asn Ala Gly
 20 25 30
 Lys Thr Thr Leu Leu His Met Leu Lys Asp Glu Arg Leu Val Gln His
 35 40 45
 Gln Pro Thr Gln Tyr Pro Thr Ser Glu Glu Leu Ser Ile Gly Arg Ile
 50 55 60
 Lys Phe Xaa Ala Phe Asp Leu Xaa Gly His Gln Ile Ala Arg Arg Val
 65 70 75 80
 Trp Lys Asp Tyr Xaa Ala
 85

<210> 832
 <211> 131
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 129
 <223> Xaa = Leu, Val

<400> 832
 Met Ser Pro Leu Glu Cys Ser Glu Cys Phe Gly Asp Gln Leu Leu His
 1 5 10 15
 Arg Thr Tyr Thr Trp Gln Leu Thr Leu His Ser Arg Pro Asn Tyr Thr
 20 25 30
 Arg Lys Arg Asp Thr Arg Ser Glu Ser Leu Glu Ile Pro Ile Ser Val
 35 40 45
 Val Leu Pro Gln Arg Gly Thr Ala Glu Pro Phe Pro Arg Leu His Asn
 50 55 60
 Leu Tyr Ser Thr Pro Arg Cys Ala Gln Gln Ala Ala Leu Pro Arg Leu
 65 70 75 80
 Ser Arg Arg Met Ala Ser Gln His Ser Tyr Pro Leu Asn Arg Phe Ser
 85 90 95
 Ser Val Pro Leu Asp Pro Met Glu Arg Pro Met Ser Gln Ala Asp Leu
 100 105 110
 Glu Leu Asp Tyr Asn Pro Pro Arg Val Gln Leu Ser Asp Glu Met Phe
 115 120 125
 Xaa Phe Gln
 130

<210> 833
 <211> 99
 <212> PRT
 <213> Homo sapiens

<220>

<221> UNSURE
 <222> 32
 <223> Xaa = * ,Lys,Leu,Met

<220>
 <221> UNSURE
 <222> 26
 <223> Xaa = Asp,Glu

<220>
 <221> UNSURE
 <222> 70,78
 <223> Xaa = His,Gln

<220>
 <221> UNSURE
 <222> 27
 <223> Xaa = Leu,Val

<400> 833
 Met Ala Thr Val Met Ala Ala Thr Ala Ala Glu Arg Ala Val Leu Glu
 1 5 10 15
 Glu Glu Phe Arg Trp Leu Leu His Asp Xaa Xaa His Ala Val Leu Xaa
 20 25 30
 Gln Leu Gln Asp Ile Leu Lys Glu Ala Ser Leu Arg Phe Thr Leu Pro
 35 40 45
 Gly Ser Gly Thr Glu Gly Pro Ala Lys Gln Glu Asn Phe Ile Leu Gly
 50 55 60
 Ser Cys Gly Thr Asp Xaa Val Lys Gly Val Leu Thr Leu Xaa Gly Asp
 65 70 75 80
 Ala Leu Ser Gln Ala Asp Val Asn Leu Lys Met Pro Arg Asn Asn Gln
 85 90 95
 Leu Leu His

<210> 834
 <211> 96
 <212> PRT
 <213> Homo sapiens

<400> 834
 Met Val Ser His Thr Phe His Met Arg Thr Glu Glu Ser Asp Ala Ser
 1 5 10 15
 Gln Glu Gly Asp Asp Leu Pro Lys Ser Ser Ala Asn Thr Ser His Pro
 20 25 30
 Lys Gln Asp Asp Ser Pro Lys Ser Ser Glu Glu Thr Ile Gln Pro Lys
 35 40 45
 Glu Gly Asp Ile Pro Lys Ala Pro Glu Glu Thr Ile Gln Ser Lys Lys
 50 55 60
 Glu Asp Leu Pro Lys Ser Ser Glu Lys Ala Ile Gln Pro Lys Glu Ser
 65 70 75 80
 Asn Ile Pro Lys Ser Ser Ala Lys Pro Ile Gln Pro Lys Leu Ala Ile
 85 90 95

<210> 835
 <211> 97
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 94
 <223> Xaa = Leu,Met

<400> 835

```

Met Ser Leu Leu Asp Gly Leu Ala Ser Ser Pro Arg Ala Pro Leu Gln
1          5          10          15
Ser Ser Lys Ala Arg Met Lys Lys Leu Pro Lys Lys Ser Gln Asn Glu
20          25          30
Lys Tyr Arg Leu Lys Tyr Leu Arg Leu Arg Lys Ala Ala Lys Ala Thr
35          40          45
Val Phe Glu Asn Ala Ala Ile Cys Asp Glu Ile Ala Arg Leu Glu Glu
50          55          60
Lys Phe Leu Lys Ala Lys Glu Glu Arg Arg Tyr Leu Leu Lys Lys Leu
65          70          75          80
Leu Gln Leu Gln Ala Leu Thr Glu Gly Val Ser Thr Gly Xaa Gln Leu
85          90          95
Leu

```

<210> 836

<211> 107

<212> PRT

<213> Homo sapiens

<400> 836

```

Met Ala Gln Glu Val Ser Glu Tyr Leu Ser Gln Asn Pro Arg Val Ala
1          5          10          15
Ala Trp Val Glu Ala Leu Arg Cys Asp Gly Glu Thr Asp Lys His Trp
20          25          30
Arg His Arg Arg Asp Phe Leu Leu Arg Asn Ala Gly Asp Leu Ala Pro
35          40          45
Ala Gly Gly Ala Ala Ser Ala Ser Thr Asp Glu Ala Ala Asp Ala Glu
50          55          60
Ser Gly Thr Arg Asn Arg Gln Leu Gln Gln Leu Ile Ser Phe Ser Met
65          70          75          80
Ala Trp Ala Asn His Val Phe Leu Gly Cys Arg Tyr Pro Gln Lys Val
85          90          95
Met Asp Lys Ile Leu Ser Met Ala Glu Gly Ile
100          105

```

<210> 837

<211> 65

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 20

<223> Xaa = Pro,Arg

<400> 837

```

Met Arg Leu Leu Gln Pro Arg Ser Arg Leu Gln Arg Thr Ser Gln Lys
1          5          10          15
Trp Leu Phe Xaa Leu His Gly Asp Glu Ser Leu Ala Pro Lys His Pro
20          25          30
Gly Ser Arg Lys Asn Met Asp Cys Tyr Trp Gln Asn Thr Ser Val His
35          40          45
Cys Arg Arg Thr Ser Leu Trp Asn Leu His Leu Pro Gly Lys Thr Leu
50          55          60
Gly
65

```

<210> 838

<211> 81

<212> PRT

<213> Homo sapiens

<220>
 <221> UNSURE
 <222> 10
 <223> Xaa = * ,Trp

<220>
 <221> UNSURE
 <222> 9
 <223> Xaa = Ala,Pro

<220>
 <221> UNSURE
 <222> 16
 <223> Xaa = Arg,Ser

<220>
 <221> UNSURE
 <222> 12
 <223> Xaa = Gly,Trp

<220>
 <221> UNSURE
 <222> 8
 <223> Xaa = Lys,Arg

<400> 838
 Met Leu Ala Arg Gly Leu Val Xaa Xaa Xaa Asn Xaa Gly Asn Arg Xaa
 1 5 10 15
 Pro Val Phe Thr Ile Ser Phe His His Leu Pro Thr Trp Asn Leu Tyr
 20 25 30
 Ile Ser Pro Ser Leu Ser Phe Leu Ile Cys Glu Thr Ala Ala Asn Thr
 35 40 45
 Gly Phe Gln Arg Val Val Lys Asn Pro Cys Glu His Leu Leu Gln Leu
 50 55 60
 Pro Ser Thr Gln Asn Arg Leu Phe His Trp Arg Arg Arg Arg Asp Gly
 65 70 75 80
 Arg

<210> 839
 <211> 50
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 47
 <223> Xaa = Phe,Ser

<400> 839
 Met Val Arg Pro Arg Arg Ala Pro Tyr Arg Ser Gly Ala Gly Gly Pro
 1 5 10 15
 Leu Gly Gly Arg Gly Arg Pro Pro Arg Pro Leu Val Val Arg Ala Val
 20 25 30
 Arg Ser Arg Ser Trp Pro Ala Ser Pro Pro Arg Pro Ala Ala Xaa Ala
 35 40 45
 Asp Pro
 50

<210> 840
 <211> 85
 <212> PRT
 <213> Homo sapiens

<400> 840

```

Met Pro Ser Phe Val Asp Arg Ile Phe Gly Gly Glu Leu Thr Ser Met
1          5          10          15
Ile Met Tyr Asp Gln Cys Arg Thr Val Ser Leu Val His Glu Ser Phe
          20          25          30
Leu Asp Leu Ser Leu Pro Val Leu Asp Asp Gln Ser Gly Lys Lys Ser
          35          40          45
Val Asn Asp Lys Asn Leu Lys Lys Thr Val Glu Asp Glu Asp Gln Asp
          50          55          60
Ser Glu Glu Glu Lys Asp Asn Asp Ser Tyr Ile Lys Glu Arg Ser Asp
65          70          75          80
Ile Pro Ser Gly Thr
          85

```

<210> 841

<211> 66

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 59

<223> Xaa = Lys,Asn

<400> 841

```

Met Ala Ser Ile Phe Ser Lys Leu Leu Thr Gly Arg Asn Ala Ser Leu
1          5          10          15
Leu Phe Ala Thr Met Gly Thr Ser Val Leu Thr Thr Gly Tyr Leu Leu
          20          25          30
Asn Arg Gln Lys Val Cys Ala Glu Val Arg Glu Gln Pro Arg Leu Phe
          35          40          45
Pro Pro Ser Ala Asp Tyr Pro Asp Leu Arg Xaa Thr Thr Thr Ala Trp
          50          55          60
Pro Ser
65

```

<210> 842

<211> 74

<212> PRT

<213> Homo sapiens

<400> 842

```

Met Ala Ala Ala Ala Gly Leu Gly Leu Gln Arg Glu Ala Ala Glu Gly
1          5          10          15
Lys Trp Ser Ala Trp Arg Met Thr Glu Glu Thr Arg Ile Val Tyr Trp
          20          25          30
Ile Lys Asp Arg Gln Leu Thr Asn Arg Asp Ser Thr Ile Leu Glu Leu
          35          40          45
Gln Lys Val Leu Lys Thr Cys Trp Phe Val Cys Trp Ser Ile Phe Gln
          50          55          60
Leu Lys Phe Leu Phe Leu Phe Phe Phe Phe
65          70

```

<210> 843

<211> 68

<212> PRT

<213> Homo sapiens

<400> 843

```

Met Lys Lys Asn Ile Trp Leu Leu Ser Gly Ser Arg Gln Gln Tyr Leu
1          5          10          15
Ile Ala Leu Trp Arg Leu Leu Leu Leu Glu Ile Tyr Ile Ser Glu Gly
          20          25          30

```

429

Thr His Ser Thr Glu Glu Asp Ile Asn Lys Gln Ile Asn Asp Lys Glu
 35 40 45
 Arg Val Ala Ala Ala Met Glu Asn Pro Asn Leu Arg Glu Ile Val Asp
 50 55 60
 Ser Val Ser Leu
 65

<210> 844
 <211> 53
 <212> PRT
 <213> Homo sapiens

<400> 844
 Met Phe Pro Pro Pro Ala Ala Ser Leu Cys Ser His Val Gly Cys Ser
 1 5 10 15
 Arg Asp His Pro Pro Val Gly Pro Gly Gln Thr Asn Arg Lys Gly Pro
 20 25 30
 Ile Thr Leu Thr Thr Ala Glu Leu Thr Ser Thr Thr Val Ser Val Ala
 35 40 45
 Leu Ala Trp Thr Pro
 50

<210> 845
 <211> 73
 <212> PRT
 <213> Homo sapiens

<400> 845
 Met Cys Pro Gly Asn Trp Leu Trp Ala Ser Met Thr Phe Met Ala Arg
 1 5 10 15
 Phe Ser Arg Ser Ser Ser Arg Ser Pro Val Arg Thr Arg Gly Thr Leu
 20 25 30
 Glu Glu Met Pro Thr Val Gln His Pro Phe Leu Asn Val Phe Glu Leu
 35 40 45
 Glu Arg Leu Leu Tyr Thr Gly Lys Thr Ala Cys Asn His Ala Asp Glu
 50 55 60
 Val Trp Pro Gly Phe Tyr Leu Gly Asp
 65 70

<210> 846
 <211> 143
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 109
 <223> Xaa = Lys,Arg

<400> 846
 Met Thr Ser Gly Ser Lys Cys Pro Ser Thr Asp Ser Gly Lys Glu Glu
 1 5 10 15
 Tyr Ile Ala Thr Phe Lys Gly Ser Glu Tyr Phe Cys Tyr Asp Leu Ser
 20 25 30
 Gln Asn Pro Ile Gln Ser Ser Ser Asp Glu Ile Thr Leu Ser Phe Lys
 35 40 45
 Thr Leu Gln Arg Asn Gly Leu Met Leu His Thr Gly Lys Ser Ala Asp
 50 55 60
 Tyr Val Asn Leu Ala Leu Lys Asn Gly Ala Val Ser Leu Val Ile Asn
 65 70 75 80
 Leu Gly Ser Gly Ala Phe Glu Ala Leu Val Glu Pro Val Asn Gly Lys
 85 90 95
 Phe Asn Asp Asn Ala Trp His Asp Val Lys Val Thr Xaa Asn Leu Arg

430

			100						105					110			
Gln	Val	Thr	Ile	Ser	Val	Asp	Gly	Ile	Leu	Thr	Thr	Thr	Gly	Tyr	Thr		
			115					120					125				
Gln	Glu	Asp	Tyr	Thr	Met	Leu	Gly	Ser	Asp	Asp	Phe	Phe	Tyr	Val			
			130				135					140					

<210> 847
 <211> 112
 <212> PRT
 <213> Homo sapiens

<400> 847

Met	Gln	Leu	Tyr	Leu	Tyr	Ser	Met	Ser	Thr	Ala	Ile	Met	Gly	Leu	Asp		
1				5					10					15			
Glu	Gln	Ala	Arg	Thr	Leu	Glu	Arg	Glu	Glu	Lys	Lys	Glu	Asn	Ser	Gly		
			20					25					30				
Ala	Ile	Arg	Ser	Asp	Arg	Cys	Val	Lys	Ala	Ser	Phe	Gln	Gly	Leu			
			35				40					45					
Ala	Asn	Asn	Ile	Ser	Phe	Tyr	Leu	Cys	Pro	Pro	Gly	Tyr	Ser	Pro	Arg		
			50			55					60						
Met	Leu	Arg	Phe	Ser	Ser	Ala	Ser	Ala	Ala	Leu	Arg	Gln	Gln	Asp	Gln		
65					70					75					80		
Pro	Leu	Leu	Leu	Leu	Phe	Leu	Leu	Leu	Ser	Leu	Leu	Asn	Val	Lys	Arg		
				85					90					95			
Met	Arg	Met	Lys	Thr	Phe	Met	Met	Thr	Arg	Ile	His	Leu	Met	Asn	Ser		
			100					105					110				

<210> 848
 <211> 54
 <212> PRT
 <213> Homo sapiens

<400> 848

Met	Ile	Gln	Arg	Asn	Leu	Ser	Asn	Asp	Asp	Asn	Ser	Ser	Pro	Ala	Cys		
1				5					10					15			
Met	Asp	Ala	Asp	Met	Asp	Ala	Val	Ser	His	Ser	Tyr	His	Trp	Met	Ser		
			20					25					30				
Arg	Asn	Ala	Arg	Cys	Asp	Pro	Ala	Ser	Leu	Tyr	Ser	Thr	Leu	Ser	Ser		
			35				40					45					
Ile	Phe	Ser	Thr	His	Ser												
			50														

<210> 849
 <211> 138
 <212> PRT
 <213> Homo sapiens

<400> 849

Met	Lys	Lys	Leu	Lys	Ala	Arg	Met	His	Gln	Ala	Ile	Glu	Arg	Phe	Tyr		
1				5					10					15			
Asp	Lys	Met	Gln	Asn	Ala	Glu	Ser	Gly	Arg	Gly	Gln	Val	Met	Ser	Ser		
			20					25					30				
Leu	Ala	Glu	Leu	Glu	Asp	Asp	Phe	Lys	Glu	Gly	Tyr	Leu	Glu	Thr	Val		
			35				40					45					
Ala	Ala	Tyr	Tyr	Glu	Glu	Gln	His	Pro	Glu	Leu	Thr	Pro	Leu	Leu	Glu		
			50			55					60						
Lys	Glu	Arg	Asp	Gly	Leu	Arg	Cys	Arg	Gly	Asn	Arg	Ser	Pro	Val	Pro		
65					70					75					80		
Asp	Val	Glu	Asp	Pro	Ala	Thr	Glu	Glu	Pro	Gly	Glu	Ser	Phe	Cys	Asp		
				85					90					95			
Lys	Val	Met	Arg	Trp	Phe	Gln	Ala	Met	Leu	Gln	Arg	Leu	Gln	Thr	Trp		
			100					105					110				
Trp	His	Gly	Val	Leu	Ala	Trp	Val	Lys	Glu	Lys	Val	Val	Ala	Leu	Val		

431

115 120 125
 His Ala Val Gln Ala Leu Trp Lys Gln Phe
 130 135

<210> 850
 <211> 70
 <212> PRT
 <213> Homo sapiens

<400> 850
 Met Ile Cys Thr Phe Leu Arg Ala Val Gln Tyr Thr Glu Lys Leu His
 1 5 10 15
 Arg Ser Ser Ala Lys Arg Leu Leu Leu Pro Tyr Ile Val Leu Asn Lys
 20 25 30
 Ala Cys Leu Lys Thr Glu Pro Ser Leu Arg Cys Gly Leu Gln Tyr Gln
 35 40 45
 Lys Lys Thr Leu Arg Pro Arg Cys Ile Leu Gly Val Thr Gln Lys Thr
 50 55 60
 Ile Trp Thr Gln Gly Pro
 65 70

<210> 851
 <211> 152
 <212> PRT
 <213> Homo sapiens

<400> 851
 Met Asn Val Thr Pro Glu Val Lys Ser Arg Gly Met Lys Phe Ala Glu
 1 5 10 15
 Glu Gln Leu Leu Lys His Gly Trp Thr Gln Gly Lys Gly Leu Gly Arg
 20 25 30
 Lys Glu Asn Gly Ile Thr Gln Ala Leu Arg Val Thr Leu Lys Gln Asp
 35 40 45
 Thr His Gly Val Gly His Asp Pro Ala Lys Glu Phe Thr Asn His Trp
 50 55 60
 Trp Asn Glu Leu Phe Asn Lys Thr Ala Ala Asn Leu Val Val Glu Thr
 65 70 75 80
 Gly Gln Asp Gly Val Gln Ile Arg Ser Leu Ser Lys Glu Thr Thr Arg
 85 90 95
 Tyr Asn His Pro Lys Pro Asn Leu Leu Tyr Gln Lys Phe Val Lys Met
 100 105 110
 Ala Thr Leu Thr Ser Gly Gly Glu Lys Pro Asn Lys Asp Leu Glu Ser
 115 120 125
 Cys Ser Asp Asp Asp Asn Gln Gly Ser Ser Pro Gln Arg Phe Leu Leu
 130 135 140
 Met Arg Cys Cys Ser Lys Pro Val
 145 150

<210> 852
 <211> 120
 <212> PRT
 <213> Homo sapiens

<400> 852
 Met Ala Lys Glu Glu Pro Gln Ser Ile Ser Arg Asp Leu Gln Glu Leu
 1 5 10 15
 Gln Lys Lys Leu Ser Leu Leu Ile Asp Ser Phe Gln Asn Asn Ser Lys
 20 25 30
 Val Val Ala Phe Met Lys Ser Pro Val Gly Gln Tyr Leu Asp Ser His
 35 40 45
 Pro Phe Leu Ala Phe Thr Leu Leu Val Phe Ile Val Met Ser Ala Val
 50 55 60
 Pro Val Gly Phe Phe Leu Leu Ile Val Val Leu Thr Thr Leu Ala Ala

432

65					70					75				80
Leu	Leu	Gly	Val	Ile	Ile	Leu	Glu	Gly	Leu	Val	Ile	Ser	Val	Gly Gly
				85					90					95
Phe	Ser	Leu	Leu	Cys	Ile	Leu	Cys	Gly	Leu	Ser	Phe	Val	Ser	Leu Ala
			100					105					110	
Met	Ser	Gly	Met	Met	Ile	Ala	Ser							
		115					120							

<210> 853
 <211> 103
 <212> PRT
 <213> Homo sapiens

<400> 853
 Met Asp Gly Thr Val Leu Cys Lys Leu Ile Asn Ser Leu Tyr Pro Pro
 1 5 10 15
 Gly Gln Glu Pro Ile Pro Lys Ile Ser Glu Ser Lys Met Ala Phe Lys
 20 25 30
 Gln Met Glu Gln Ile Ser Gln Phe Leu Lys Ala Ala Glu Thr Tyr Gly
 35 40 45
 Val Arg Thr Thr Asp Ile Phe Gln Thr Val Asp Leu Trp Glu Gly Lys
 50 55 60
 Asp Met Ala Ala Val Gln Arg Thr Leu Met Ala Leu Gly Ser Val Ala
 65 70 75 80
 Val Thr Lys Asp Asp Gly Cys Tyr Arg Gly Glu Pro Ser Trp Phe His
 85 90 95
 Arg Lys Ala Gln Gln Asn Arg
 100

<210> 854
 <211> 78
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 68
 <223> Xaa = Gly,Arg

<220>
 <221> UNSURE
 <222> 76
 <223> Xaa = Lys,Asn

<400> 854
 Met Pro Leu Pro Arg Lys Asp Asn Pro Gln Cys Ser Pro Leu Leu Pro
 1 5 10 15
 Pro Leu Ala Ser Val Gly Lys Lys Met Ala Lys Lys Ile Ala Ala Ala
 20 25 30
 Pro Pro Pro Pro Pro Pro Pro Ser Glu Arg Arg Gly Lys Arg Ala
 35 40 45
 Glu Gln Lys Gly Ala Val Gly Gln Arg Pro Asp Trp Ile Phe Ser Pro
 50 55 60
 Gln Pro Gln Xaa Met Phe Leu Met Thr Gln Gly Xaa Leu Val
 65 70 75

<210> 855
 <211> 136
 <212> PRT
 <213> Homo sapiens

<400> 855
 Met Ser Glu Pro Gly Lys Leu Ser Gln Lys Ile Lys Val Trp Leu Gln

433

```

1           5           10           15
Glu Tyr Trp Asn Ile Thr Asp Leu Val Ala Ile Ser Thr Phe Met Ile
      20      25      30
Gly Ala Ile Leu Arg Leu Gln Asn Gln Pro Tyr Met Gly Tyr Gly Arg
      35      40      45
Val Ile Tyr Cys Val Asp Ile Ile Phe Trp Tyr Ile Arg Val Leu Asp
      50      55      60
Ile Phe Gly Val Asn Lys Tyr Leu Gly Pro Tyr Val Met Met Ile Gly
65      70      75      80
Lys Met Met Ile Asp Met Leu Tyr Phe Val Val Ile Met Leu Val Val
      85      90      95
Leu Met Ser Phe Gly Val Ala Arg Gln Ala Ile Leu His Pro Glu Glu
      100      105      110
Lys Pro Ser Trp Lys Leu Ala Arg Asn Ile Phe Tyr Met Pro Tyr Trp
      115      120      125
Met Ile Tyr Gly Glu Val Leu Gln
      130      135

```

<210> 856
 <211> 118
 <212> PRT
 <213> Homo sapiens

```

<400> 856
Met Val Arg Gly Trp Ala Pro Thr Trp Ala Ile Pro Glu Leu Cys Gly
1           5           10           15
Val Trp Thr Leu Thr Gly Thr Pro Ser Met Ser Ser Leu Ala Gln Leu
      20      25      30
Thr Thr Ala Val Val Ser Gly Thr Val Lys Gln Ser Gly Glu Val Leu
      35      40      45
Val Asn Val Lys Glu His Ser Arg Gln Ile Asn Asp Ile Gln Leu Ser
      50      55      60
Arg Asp Met Thr Met Phe Val Thr Ala Ser Lys Asp Asn Thr Ala Lys
65      70      75      80
Leu Phe Asp Ser Thr Thr Leu Glu His Gln Lys Thr Phe Arg Thr Glu
      85      90      95
Arg Pro Val Asn Ser Ala Ala Leu Ser Pro Asn Tyr Asp His Val Val
      100      105      110
Leu Gly Gly Gly Gln Glu
      115

```

<210> 857
 <211> 122
 <212> PRT
 <213> Homo sapiens

```

<400> 857
Met Arg Pro Ser Cys Ala Pro Gly Trp Phe Tyr His Lys Ser Asn Cys
1           5           10           15
Tyr Gly Tyr Phe Arg Lys Leu Arg Asn Trp Ser Asp Ala Glu Leu Glu
      20      25      30
Cys Gln Ser Tyr Gly Asn Gly Ala His Leu Ala Ser Ile Leu Ser Leu
      35      40      45
Lys Glu Ala Ser Thr Ile Ala Glu Tyr Ile Ser Gly Tyr Gln Arg Ser
      50      55      60
Gln Pro Ile Trp Ile Gly Leu His Asp Pro Gln Lys Arg Gln Gln Trp
65      70      75      80
Gln Trp Ile Asp Gly Ala Met Tyr Leu Tyr Arg Ser Trp Ser Gly Lys
      85      90      95
Ser Met Gly Gly Asn Lys His Cys Ala Glu Met Ser Ser Asn Asn Asn
      100      105      110
Phe Leu Thr Trp Ser Ser Asn Glu Cys Asn
      115      120

```

<210> 858
 <211> 107
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -23..-1

<400> 858
 Met Leu Leu Gln Ser Gly Leu Ser Phe Arg Arg Leu Leu Leu Leu Ser
 -20 -15 -10
 Leu Val Ser Gly Ala Leu Gly Leu Gly Gly Ala Val Leu Gly Val Gly
 -5 1 5
 Leu Ser Leu Gly Pro Val Pro Leu Thr Pro Trp Val Phe Gly Val Thr
 10 15 20 25
 Ala Gly Val Phe Leu Tyr Val Ala Leu Val Asp Met Leu Pro Ala Leu
 30 35 40
 Leu Arg Pro Pro Glu Pro Leu Pro Thr Pro His Val Leu Leu Gln Gly
 45 50 55
 Leu Gly Leu Leu Leu Gly Gly Gly Leu Met Leu Ala Ile Thr Leu Leu
 60 65 70
 Glu Glu Arg Leu Leu Pro Val Thr Thr Glu Gly
 75 80

<210> 859
 <211> 92
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -24..-1

<220>
 <221> UNSURE
 <222> 52
 <223> Xaa = Arg,Thr

<400> 859
 Met Val Met Gly Leu Gly Val Leu Leu Leu Val Phe Val Leu Gly Leu
 -20 -15 -10
 Gly Leu Thr Pro Thr Leu Ala Gln Asp Asn Ser Arg Tyr Thr His
 -5 1 5
 Phe Leu Thr Gln His Tyr Asp Ala Lys Pro Gln Gly Arg Asp Asp Arg
 10 15 20
 Tyr Cys Glu Ser Ile Met Arg Arg Arg Gly Leu Thr Ser Pro Cys Lys
 25 30 35 40
 Asp Ile Asn Thr Phe Ile His Gly Asn Lys Arg Xaa Ser Arg Pro Ser
 45 50 55
 Val Lys Thr Arg Met Glu Thr Leu Thr Glu Lys Thr
 60 65

<210> 860
 <211> 80
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -32..-1

<220>

<221> UNSURE

<222> 44

<223> Xaa = Ala,Pro

<400> 860

Met	Met	Leu	Val	Gly	Phe	Leu	Gly	Cys	Cys	Gly	Ala	Val	Gln	Glu	Ser
		-30					-25					-20			
Gln	Cys	Met	Leu	Gly	Leu	Phe	Phe	Gly	Phe	Leu	Leu	Val	Ile	Phe	Ala
	-15					-10					-5				
Ile	Glu	Ile	Ala	Ala	Ala	Ile	Trp	Gly	Tyr	Ser	His	Lys	Asp	Glu	Val
1			5						10					15	
Ile	Lys	Glu	Val	Gln	Glu	Phe	Tyr	Lys	Asp	Thr	Tyr	Asn	Lys	Leu	Lys
		20						25				30			
Thr	Lys	Asp	Glu	Pro	Gln	Arg	Glu	Thr	Leu	Lys	Xaa	Ser	Thr	Met	Arg
		35					40					45			

<210> 861

<211> 166

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -15..-1

<220>

<221> UNSURE

<222> 142

<223> Xaa = Ser,Thr

<400> 861

Met	Leu	Pro	Leu	Leu	Ile	Ile	Cys	Leu	Leu	Pro	Ala	Ile	Glu	Gly	Lys
-15					-10					-5					1
Asn	Cys	Leu	Arg	Cys	Trp	Pro	Glu	Leu	Ser	Ala	Leu	Ile	Asp	Tyr	Asp
		5					10					15			
Leu	Gln	Ile	Leu	Trp	Val	Thr	Pro	Gly	Pro	Pro	Thr	Glu	Leu	Ser	Gln
	20						25				30				
Asn	Arg	Asp	His	Leu	Glu	Glu	Glu	Thr	Ala	Lys	Phe	Phe	Thr	Gln	Val
	35				40					45					
His	Gln	Ala	Ile	Lys	Thr	Leu	Arg	Asp	Asp	Lys	Thr	Val	Leu	Leu	Glu
50					55					60					65
Glu	Ile	Tyr	Thr	His	Lys	Asn	Leu	Phe	Thr	Glu	Arg	Leu	Asn	Lys	Ile
			70						75					80	
Ser	Asp	Gly	Leu	Lys	Glu	Lys	Asp	Ile	Gln	Ser	Thr	Leu	Lys	Val	Thr
		85					90					95			
Ser	Cys	Ala	Asp	Cys	Arg	Thr	His	Phe	Leu	Ser	Cys	Asn	Asp	Pro	Thr
	100					105					110				
Phe	Cys	Pro	Ala	Arg	Asn	Arg	Arg	Thr	Ser	Leu	Trp	Ala	Val	Ser	Leu
	115				120					125					
Ser	Ser	Ala	Leu	Leu	Leu	Ala	Ile	Ala	Gly	Asp	Val	Xaa	Phe	Thr	Gly
130					135					140					145
Lys	Gly	Arg	Arg	Arg	Gln										
					150										

<210> 862

<211> 54

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -26..-1

<400> 862

```

Met Val Lys Val Met Trp Val Ser Val Leu Ala Leu Ala Ala Ala Ile
  -25                -20                -15
Leu Leu Leu Thr Val Pro Val Ala Glu Gly Val Thr Cys Ser Pro Met
-10                -5                1                5
Gln Leu Ala Ser Cys Ala Ala Ala Met Thr Ser Ser Ser Pro Pro Ser
          10                15                20
Glu Ala Val Ala Gln Ser
      25

```

<210> 863

<211> 55

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -26..-1

<400> 863

```

Met Lys Pro Val Ser Arg Arg Thr Leu Asp Trp Ile Tyr Ser Val Leu
  -25                -20                -15
Leu Leu Ala Ile Val Leu Ile Ser Trp Gly Cys Ile Ile Tyr Ala Ser
-10                -5                1                5
Met Val Ser Ala Arg Arg Gln Leu Arg Lys Lys Tyr Pro Asp Lys Ile
          10                15                20
Phe Gly Thr Asn Glu Asn Leu
      25

```

<210> 864

<211> 101

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -22..-1

<400> 864

```

Met Arg Thr Leu Ala Ile Leu Ala Ala Ile Leu Leu Val Ala Leu Gln
  -20                -15                -10
Ala Arg Leu Ser His Ser Arg Gln Glu Leu Met Arg Leu Leu Gln Pro
  -5                1                5                10
Arg Ser Arg Leu Gln Arg Thr Ser Gln Lys Trp Leu Phe Pro Leu His
          15                20                25
Gly Thr Lys Ala Trp Leu Gln Ser Ile Gln Ala Gln Gly Lys Thr Trp
          30                35                40
Pro Ala Ile Ala Glu Tyr Gln Arg Ala Leu Gln Glu Asn Val Ala Met
          45                50                55
Glu Pro Ala Ser Thr Arg Glu Asp Ser Gly His Ser Ala Ala Glu Leu
          60                65                70
Ala Glu Lys Glu Lys
      75

```

<210> 865

<211> 57

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -21..-1

<400> 865

```

Met Val Leu Leu Ser Ser Thr Met Ser Val Leu Ile Phe Cys Leu Leu
  -20                      -15          -10
Asn Leu Ala Ile Ser Asp Arg Arg Val Met Lys Ser Pro Thr Ile Ile
  -5                      1          5          10
Val Ile His Leu Phe Phe Pro His Ser Ser Thr Ser Ile Cys Ile Glu
          15                      20          25
Asn Leu Val Thr Leu Leu Leu Gly Thr
      30                      35

```

<210> 866

<211> 129

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -26..-1

<400> 866

```

Met Val Ser Arg Met Val Ser Thr Met Leu Ser Gly Leu Leu Phe Trp
  -25                      -20          -15
Leu Ala Ser Gly Trp Thr Pro Ala Phe Ala Tyr Ser Pro Arg Thr Pro
  -10                      -5          1          5
Asp Arg Val Ser Glu Ala Asp Ile Gln Arg Leu Leu His Gly Val Met
          10                      15          20
Glu Gln Leu Gly Ile Ala Arg Pro Arg Val Glu Tyr Pro Ala His Gln
          25                      30          35
Ala Met Asn Leu Val Gly Pro Gln Ser Ile Glu Gly Gly Ala His Glu
          40                      45          50
Gly Leu Gln His Leu Gly Pro Phe Gly Asn Ile Pro Asn Ile Val Ala
          55                      60          65          70
Glu Leu Thr Gly Asp Asn Ile Pro Lys Asp Phe Ser Glu Asp Gln Gly
          75                      80          85
Tyr Gln Thr Leu Gln Ile Pro Val Leu Leu Glu Lys Gln Met Met Asp
          90                      95          100
Val

```

<210> 867

<211> 99

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -41..-1

<220>

<221> UNSURE

<222> -36

<223> Xaa = Ala,Cys,Gly,Ser,Trp

<400> 867

```

Met Ala Ala Ala Val Xaa Ala Ala Pro Gly Ala Leu Gly Ser Leu His
  -40                      -35          -30
Ala Gly Gly Ala Arg Leu Val Ala Ala Cys Ser Ala Trp Leu Cys Pro
  -25                      -20          -15          -10
Gly Leu Arg Leu Pro Gly Ser Leu Ala Gly Arg Arg Ala Gly Pro Ala
          -5                      1          5
Ile Trp Ala Gln Gly Trp Val Pro Ala Ala Gly Gly Pro Ala Pro Lys
          10                      15          20
Arg Gly Tyr Ser Ser Glu Met Lys Thr Glu Asp Glu Leu Arg Val Arg

```

438

25 30 35
 His Leu Glu Glu Glu Asn Arg Asp Arg Asp Lys Asn Met Val Lys Gly
 40 45 50 55
 Met Glu Met

<210> 868
 <211> 108
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -19..-1

<400> 868
 Met Ala Ala Ala Thr Gly Ala Val Ala Ala Ser Ala Ala Ser Gly Gln
 -15 -10 -5
 Ala Glu Gly Lys Lys Ile Thr Asp Leu Arg Val Ile Asp Leu Lys Ser
 1 5 10
 Glu Leu Lys Arg Arg Asn Leu Asp Ile Thr Gly Val Lys Thr Val Leu
 15 20 25
 Ile Ser Arg Leu Lys Gln Ala Ile Glu Glu Glu Gly Gly Asp Pro Asp
 30 35 40 45
 Asn Ile Glu Leu Thr Val Ser Thr Asp Thr Pro Asn Lys Lys Pro Thr
 50 55 60
 Lys Gly Lys Gly Lys Lys His Glu Ala Asp Glu Leu Ser Gly Asp Ala
 65 70 75
 Ser Val Glu Asp Asp Ala Phe Ile Lys Val Lys Cys
 80 85

<210> 869
 <211> 60
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -41..-1

<400> 869
 Met Pro Glu Ser Leu Trp Gln His Ser His Trp Cys Gly Glu Ser Phe
 -40 -35 -30
 Gln Leu Lys Glu Ala Gly Phe Leu Gly Thr Leu Glu Val Phe Trp Met
 -25 -20 -15 -10
 Ser Phe Tyr Leu Ser Arg Val Asn Cys Thr Thr Lys Glu Thr Ser Pro
 -5 1 5
 Leu Pro Lys Pro Ala Arg His Leu Gly Leu Glu Pro
 10 15

<210> 870
 <211> 60
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -45..-1

<400> 870
 Met His Ser Leu Phe Ile Ala Ser Leu Lys Val Leu Phe Tyr Tyr Ser
 -45 -40 -35 -30
 Phe Ser Phe Arg Phe Asn Trp Phe Asp Cys Leu Leu His Asn Leu Gly
 -25 -20 -15

439

Glu Asn Phe Leu Ser Leu Leu Ser Lys Ser Cys Ser Ala Asp Pro Ser
 -10 -5 1
 Gly Ser Thr Phe Met Arg Asp Ile Glu Thr Asn Lys
 5 10 15

<210> 871
 <211> 133
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -13..-1

<400> 871
 Met Ala His Cys Val Thr Leu Val Gln Leu Ser Ile Ser Cys Asp His
 -10 -5 1
 Leu Ile Asp Lys Asp Ile Gly Ser Lys Ser Asp Pro Leu Cys Val Leu
 5 10 15
 Leu Gln Asp Val Gly Gly Gly Ser Trp Ala Glu Leu Gly Arg Thr Glu
 20 25 30 35
 Arg Val Arg Asn Cys Ser Ser Pro Glu Phe Ser Lys Thr Leu Gln Leu
 40 45 50
 Glu Tyr Arg Phe Glu Thr Val Gln Lys Leu Arg Phe Gly Ile Tyr Asp
 55 60 65
 Ile Asp Asn Lys Thr Pro Glu Leu Arg Asp Asp Asp Phe Leu Gly Gly
 70 75 80
 Ala Glu Cys Ser Leu Gly Gln Ile Val Ser Ser Gln Val Leu Thr Leu
 85 90 95
 Pro Leu Met Leu Lys Pro Gly Lys Pro Ala Gly Arg Gly Thr Ser Arg
 100 105 110 115
 Ser Gln Leu Arg Asn
 120

<210> 872
 <211> 68
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -47..-1

<400> 872
 Met Val Lys Ser Ser Leu Gln Arg Ile Leu Asn Ser His Cys Phe Ala
 -45 -40 -35
 Arg Glu Lys Glu Gly Asp Lys Pro Ser Ala Thr Ile His Ala Ser Arg
 -30 -25 -20
 Thr Met Pro Leu Leu Ser Leu His Ser Arg Gly Gly Ser Ser Ser Glu
 -15 -10 -5 1
 Ser Ser Arg Val Ser Leu His Cys Cys Ser Asn Pro Gly Pro Gly Pro
 5 10 15
 Arg Trp Cys Ser
 20

<210> 873
 <211> 67
 <212> PRT
 <213> Homo sapiens

<220>
 <221> SIGNAL
 <222> -31..-1

<400> 873

```

Met Ser Arg Leu Gly Ser Ala Val Pro Asp Cys His Ile Asn Ala Pro
  -30                      -25                -20
Glu Pro Ala Ser Trp Ala Leu Gly Arg Trp Gly Ser Cys Arg Ala Leu
-15                      -10                -5                1
Asp Lys Gly Glu Gly Gly Arg Glu Gly Asp Pro Lys His Pro Ser Leu
      5                      10                15
Pro Gly Pro Arg Arg Gln Thr Cys Pro Met Arg Ser Thr Leu Met Asn
      20                      25                30
Cys Trp Ser
      35

```

<210> 874

<211> 81

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -19..-1

<400> 874

```

Met Ala Ala Ala Val Gly Arg Leu Leu Arg Ala Ser Val Ala Arg His
      -15                      -10                -5
Val Ser Ala Ile Pro Trp Gly Ile Ser Ala Thr Ala Ala Leu Arg Pro
      1                      5                10
Ala Ala Cys Gly Arg Thr Ser Leu Thr Asn Leu Leu Cys Ser Gly Ser
      15                      20                25
Ser Gln Ala Lys Ile Ile Ser Ser Ile Leu Trp Ile Ser Pro Leu Cys
30                      35                40                45
Val Leu Gln Lys Leu Leu Leu Leu Val Thr Lys Leu Thr Asn Phe Thr
      50                      55                60
Met

```

<210> 875

<211> 53

<212> PRT

<213> Homo sapiens

<220>

<221> SIGNAL

<222> -35..-1

<400> 875

```

Met Glu Gly Arg Ser Arg Ser Arg Lys Ala Cys Leu Glu Val Ala Thr
-35                      -30                -25                -20
Val Asn Gln Val Phe Val Gly Leu Val Glu Val Val Ala Val Gly Val
      -15                      -10                -5
Val Arg Ser Asp Gln Ile Lys Asn Met Tyr Leu Lys Tyr Ser Gln His
      1                      5                10
Asp Leu Leu Met Asp
      15

```

<210> 876

<211> 56

<212> PRT

<213> Homo sapiens

<400> 876

```

Met Gln Glu Met Ser Arg Pro Ile Asn Lys Lys Phe Leu Ser Leu Ala
1                      5                10                15
Glu Asn Gly Trp Pro His Glu Ala Thr Thr Thr Gly Gly Lys Glu Gly

```

441

20 25 30
 Cys Ser Thr Pro Arg Leu Ile Gly Lys Pro Asn Gln Ile Trp Asp Leu
 35 40 45
 Tyr Ser Val Ala Glu Ala Gly Arg
 50 55

<210> 877
 <211> 53
 <212> PRT
 <213> Homo sapiens

<400> 877
 Met Lys Lys Pro Gln Ser Glu Asn Phe Gly Gly Gly Pro Arg Thr Trp
 1 5 10 15
 Ile Asp Leu Leu Val Val Phe Pro Lys Val His Val Leu Glu Ile Phe
 20 25 30
 Ser Leu Ser Tyr Tyr Ser Leu Lys Lys Lys Met Leu Thr Val Tyr Leu
 35 40 45
 Phe Ser Ser Ile Leu
 50

<210> 878
 <211> 58
 <212> PRT
 <213> Homo sapiens

<400> 878
 Met Ser Thr Lys Arg Arg Leu Glu Glu Glu Gln Glu Pro Leu Arg Lys
 1 5 10 15
 Gln Phe Leu Ser Glu Glu Asn Met Ala Thr His Phe Ser Gln Leu Ser
 20 25 30
 Leu His Asn Asp Gln Pro Leu Leu Gln Pro His His Asp Leu Leu Pro
 35 40 45
 Ser Pro Ala Pro Thr Gln Glu Pro Leu Leu
 50 55

<210> 879
 <211> 61
 <212> PRT
 <213> Homo sapiens

<400> 879
 Met Ala Ser Asp Leu Asp Phe Ser Pro Pro Glu Cys Pro Ser Pro Leu
 1 5 10 15
 Ser Trp Arg Thr Cys Tyr Gly Thr Asp Ser Ser Trp Glu Pro Ser Ser
 20 25 30
 Ser Ser Ser Val Cys Trp Pro Ser Ser Tyr Pro Phe Pro Ser Pro Thr
 35 40 45
 Arg Arg Arg Leu Asn Arg Leu Ser Pro Glu Val Leu Arg
 50 55 60

<210> 880
 <211> 61
 <212> PRT
 <213> Homo sapiens

<400> 880
 Met Leu Ser Leu Asp Phe Leu Asp Asp Val Arg Arg Met Asn Lys Arg
 1 5 10 15
 Gln Leu Tyr Tyr Gln Val Leu Asn Phe Gly Met Ile Val Ser Ser Ala
 20 25 30
 Leu Met Ile Trp Lys Gly Leu Met Val Ile Thr Gly Ser Glu Ser Pro
 35 40 45

442

Ile Val Val Val Ser Gly Lys Met Gly Ile Ser Ser Phe
 50 55 60

<210> 881
 <211> 113
 <212> PRT
 <213> Homo sapiens

<400> 881
 Met Ile Val Ser Ser Ala Leu Met Ile Trp Lys Gly Leu Met Val Ile
 1 5 10 15
 Thr Gly Ser Glu Ser Pro Ile Val Val Val Leu Ser Gly Ser Met Glu
 20 25 30
 Pro Ala Phe His Arg Gly Asp Leu Leu Phe Leu Thr Asn Arg Val Glu
 35 40 45
 Asp Pro Ile Arg Val Gly Glu Ile Val Val Phe Arg Ile Glu Gly Arg
 50 55 60
 Glu Ile Pro Ile Val His Arg Val Leu Lys Ile His Glu Lys Phe Val
 65 70 75 80
 Pro Tyr Ile Gly Ile Val Thr Ile Leu Met Asn Asp Tyr Pro Lys Phe
 85 90 95
 Lys Tyr Ala Val Leu Phe Leu Leu Gly Leu Phe Val Leu Val His Arg
 100 105 110
 Glu

<210> 882
 <211> 72
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 17
 <223> Xaa = Cys,Gly,Arg,Ser

<220>
 <221> UNSURE
 <222> 61
 <223> Xaa = His,Gln

<400> 882
 Met Gly Ser Ser Arg Asp Thr Ser Arg Ser Leu Gly Gly Cys Ser Pro
 1 5 10 15
 Xaa Ser Ser Leu Asn Gly Phe Gly Phe Lys Asp Asp His Leu Arg Leu
 20 25 30
 Leu Gly Cys His Ser Ala Thr Asp His Leu Gln Leu Leu Ala Thr Leu
 35 40 45
 Gly Phe Leu His Phe Cys Leu Tyr Val Leu Leu Leu Xaa Arg Leu Ser
 50 55 60
 His Asp Ile Tyr Leu Ser Val Met
 65 70

<210> 883
 <211> 70
 <212> PRT
 <213> Homo sapiens

<400> 883
 Met Trp Tyr Ala Lys Lys Gln Gly Gln Lys Asn Lys Glu Ala Glu Arg
 1 5 10 15
 Gln Glu Arg Ala Val Val His Met Asp Glu Arg Arg Glu Glu Gln Ile
 20 25 30
 Val Gln Leu Leu Asn Ser Val Gln Ala Lys Asn Asp Lys Glu Ser Glu

443

35 40 45
 Ala Gln Ile Ser Thr Pro Asp Leu Ala Val Arg Ser Ala Pro Ser Gly
 50 55 60
 Gln Pro Phe Arg Ala Cys
 65 70

<210> 884
 <211> 53
 <212> PRT
 <213> Homo sapiens

<400> 884
 Met Ser Gly Phe Glu Asn Leu Asn Thr Asp Phe Tyr Gln Thr Ser Tyr
 1 5 10 15
 Ser Ile Asp Asp Gln Ser Gln Gln Ser Tyr Asp Tyr Gly Gly Ser Gly
 20 25 30
 Gly Pro Tyr Asn Ser Met Leu Ala Met Thr Ile Arg Ser Lys Ala Asp
 35 40 45
 Leu Ser Leu Gln Thr
 50

<210> 885
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 885
 Met Ile Phe Cys Ser His Lys Leu Ala Leu Leu Ser Ile Phe Pro Glu
 1 5 10 15
 His Thr Pro Gly Val Cys Lys Tyr Ser Leu Pro Leu Leu His Thr His
 20 25 30
 Arg Val Phe Thr Ser Val Thr Phe Cys Ala Tyr Cys Ser Ala Pro Cys
 35 40 45
 Leu Phe His Leu Thr Val Ile Ile
 50 55

<210> 886
 <211> 103
 <212> PRT
 <213> Homo sapiens

<400> 886
 Met Met Gly Ser Lys Met Ala Ser Ala Ser Arg Val Val Gln Val Val
 1 5 10 15
 Lys Pro His Thr Pro Leu Ile Arg Phe Pro Asp Arg Arg Asp Asn Pro
 20 25 30
 Lys Pro Asn Val Ser Glu Ala Leu Arg Ser Ala Gly Leu Pro Ser His
 35 40 45
 Ser Ser Val Ile Ser Gln His Ser Lys Gly Ser Lys Ser Pro Asp Leu
 50 55 60
 Leu Met Tyr Gln Gly Pro Pro Asp Thr Ala Glu Ile Ile Lys Thr Leu
 65 70 75 80
 Pro Gln Lys Tyr Arg Arg Lys Leu Val Ser Gln Glu Glu Met Glu Phe
 85 90 95
 Ile Gln Arg Gly Gly Pro Glu
 100

<210> 887
 <211> 124
 <212> PRT
 <213> Homo sapiens

<400> 887

444

Met Ala Arg Val Ile Ile Asp Asn Lys Asp Gly Pro Thr Gln Lys Tyr
 1 5 10 15
 Leu Leu Ile Phe Gly Ala Phe Val Ser Val Tyr Ile Gln Glu Met Phe
 20 25 30
 Arg Phe Ala Tyr Tyr Lys Leu Leu Lys Lys Ala Ser Glu Gly Leu Lys
 35 40 45
 Ser Ile Asn Pro Gly Glu Thr Ala Pro Ser Met Arg Leu Leu Ala Tyr
 50 55 60
 Val Ser Gly Leu Gly Phe Gly Ile Met Ser Gly Val Phe Ser Phe Val
 65 70 75 80
 Asn Thr Leu Ser Asp Ser Leu Gly Pro Gly Thr Val Gly Ile His Gly
 85 90 95
 Asp Ser Pro Gln Phe Phe Leu Leu Phe Ser Phe His Asp Ala Gly His
 100 105 110
 Tyr Leu Ala Ala Cys Ile Leu Gly His Cys Ile Phe
 115 120

<210> 888
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 888
 Met Ala Ser Ser Thr Val Pro Val Ser Ala Ala Gly Ser Ala Asn Glu
 1 5 10 15
 Thr Pro Glu Ile Pro Asp Asn Val Gly Asp Trp Leu Arg Gly Val Tyr
 20 25 30
 Pro Leu Pro Leu Ile Gly Met Thr Ser Gly Glu Leu Asp Thr Lys Phe
 35 40 45
 Gly Thr Leu Cys Cys Gly Ser Leu Ala Gly Gln Glu Leu Glu
 50 55 60

<210> 889
 <211> 65
 <212> PRT
 <213> Homo sapiens

<400> 889
 Met Leu Arg His Ser Pro Ser Leu Trp Glu Leu Val Glu Glu His Val
 1 5 10 15
 Pro Leu Arg Glu Arg Arg Glu Val Lys Arg Ile Leu Gly Gly Gly Gly
 20 25 30
 Gly Gly Pro Glu Pro Gly Ala Ala Gly Gly Gly Gly Asp Val Thr Gly
 35 40 45
 Thr Ala Pro Arg Gly Ser Ile Leu Ser Ser Pro Gln Leu Pro Pro His
 50 55 60
 Leu
 65

<210> 890
 <211> 140
 <212> PRT
 <213> Homo sapiens

<400> 890
 Met Gly Gln Asn Arg Glu Met Leu Pro Phe Trp Met Asn Ser Thr Gly
 1 5 10 15
 Arg Arg Glu Gly Trp Gln Arg Gly Trp His Gly Tyr Asp Asn Glu Leu
 20 25 30
 Met Asp Met Arg Gly Ile Phe Leu Ala Phe Gly Pro Asp Phe Lys Ser
 35 40 45
 Asn Phe Arg Ala Ala Pro Ile Arg Ser Val Asp Val Tyr Asn Val Met
 50 55 60

445

Cys Asn Val Val Gly Ile Thr Pro Leu Pro Asn Asn Gly Ser Trp Ser
 65 70 75 80
 Arg Val Met Cys Met Leu Lys Gly Arg Ala Ala Leu Pro Arg Leu Ser
 85 90 95
 Gly Pro Ala Thr Val Pro Trp His Asp Ser Ser Leu Pro Ala Cys Ile
 100 105 110
 Thr Asp His Ile Ala Cys Leu Arg Lys Lys His His Gln Gln Ser Gly
 115 120 125
 Pro Pro Lys Pro Asp Asp Phe His Phe Met Cys Glu
 130 135 140

<210> 891
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 891
 Met Glu Met Ala Ser Ser Ala Gly Ser Trp Leu Ser Gly Cys Leu Ile
 1 5 10 15
 Pro Leu Val Phe Leu Arg Leu Ser Val His Val Ser Gly His Ala Gly
 20 25 30
 Asp Ala Gly Lys Phe His Val Ala Leu Leu Gly Gly Thr Ala Glu Leu
 35 40 45
 Leu Cys Pro Leu Ser Leu Trp Pro Gly Thr Tyr Pro Arg Arg
 50 55 60

<210> 892
 <211> 114
 <212> PRT
 <213> Homo sapiens

<400> 892
 Met Asp Ile Val Ala His Cys Glu Lys Ala Met Ala Glu Ser Val Lys
 1 5 10 15
 Lys His Gln Phe Ser Thr Ser Glu Ser Tyr Leu Pro Ala Thr Leu Ile
 20 25 30
 Leu Asn Ser Ile Ser Phe Asn Thr Asp Gly Ser Ile Gln Tyr Lys Glu
 35 40 45
 Lys Leu Tyr Phe Ser Ala Ser Glu Ala Leu Asp Ala Tyr Ile Asp Asp
 50 55 60
 Phe His Leu Asn Tyr Glu Pro Pro Asp Ile Asp Thr Lys Val Asn Leu
 65 70 75 80
 Asp Gln Ser Pro Leu Glu Phe Leu Ala Lys Ser Asn Ser Gly Val His
 85 90 95
 Gly Ile Ala Pro Gly Gln Ser Gly Met Glu Lys Ile Ile Gln Trp Tyr
 100 105 110
 Ala Ala

<210> 893
 <211> 55
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 27
 <223> Xaa = Asn,Thr

<220>
 <221> UNSURE
 <222> 39
 <223> Xaa = Leu,Val

<220>
 <221> UNSURE
 <222> 34,37
 <223> Xaa = Lys,Asn

<400> 893
 Met Arg Cys Arg Pro Ser Ser Val His Val Ala Ile Leu Thr Glu Ala
 1 5 10 15
 Ile Pro Pro Lys Met Ile Ser Phe Ser Thr Xaa Ser Lys Arg Ile Phe
 20 25 30
 Lys Xaa Ser Pro Xaa Met Xaa Lys Ser Gln Val Cys Pro Leu Ile Gly
 35 40 45
 Glu Ile Ser Gly Ser Trp Met
 50 55

<210> 894
 <211> 123
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 26
 <223> Xaa = His,Gln

<400> 894
 Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg Gln
 1 5 10 15
 Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser Arg
 20 25 30
 Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala Gly
 35 40 45
 Asp Pro Lys Lys Glu Lys Lys Ser Leu Asp Ser Asp Glu Ser Glu Asp
 50 55 60
 Glu Glu Asp Asp Tyr Gln Gln Lys Arg Lys Gly Val Glu Gly Leu Ile
 65 70 75 80
 Asp Ile Glu Asn Pro Asn Arg Val Ala Gln Thr Thr Lys Lys Val Thr
 85 90 95
 Gln Leu Asp Leu Asp Gly Pro Arg Ser Phe Arg Gly Glu Asn Glu Lys
 100 105 110
 Arg Leu Arg Ser Arg Arg Gln Lys Ser Val Thr
 115 120

<210> 895
 <211> 123
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 26
 <223> Xaa = His,Gln

<400> 895
 Met Pro Lys Gly Gly Arg Lys Gly Gly His Lys Gly Arg Ala Arg Gln
 1 5 10 15
 Tyr Thr Ser Pro Glu Glu Ile Asp Ala Xaa Cys Arg Leu Arg Ser Arg
 20 25 30
 Arg Pro Gly Lys Lys Arg Ser Lys Lys Gly Gly Asp Gly Ala Ala Gly
 35 40 45
 Asp Pro Lys Lys Glu Lys Lys Ser Leu Asp Ser Asp Glu Ser Glu Asp
 50 55 60
 Glu Glu Asp Asp Tyr Gln Gln Lys Arg Lys Gly Val Glu Gly Leu Ile

447

65					70						75				80
Asp	Ile	Glu	Asn	Pro	Asn	Arg	Val	Ala	Gln	Thr	Thr	Lys	Lys	Val	Thr
				85					90					95	
Gln	Leu	Asp	Leu	Asp	Gly	Pro	Arg	Ser	Phe	Arg	Gly	Glu	Asn	Glu	Lys
			100					105					110		
Arg	Leu	Arg	Ser	Arg	Arg	Gln	Lys	Ser	Val	Thr					
		115					120								

<210> 896

<211> 81

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 78

<223> Xaa = Ser, Thr

<400> 896

Met	Val	Gly	Leu	Lys	Lys	Lys	Ser	Ala	Asp	Asp	Val	Lys	Lys	Val	Phe
1				5					10					15	
His	Met	Leu	Asp	Lys	Asp	Lys	Ser	Gly	Phe	Ile	Glu	Glu	Asp	Glu	Leu
			20					25					30		
Gly	Phe	Ile	Leu	Lys	Gly	Phe	Ser	Pro	Asp	Ala	Arg	Asp	Leu	Ser	Ala
		35				40						45			
Lys	Glu	Thr	Lys	Met	Leu	Met	Ala	Ala	Gly	Asp	Lys	Asp	Gly	Asp	Gly
	50					55					60				
Lys	Ile	Gly	Val	Asp	Glu	Phe	Ser	Thr	Leu	Val	Ala	Glu	Xaa	Lys	Lys
65					70					75					80
His															

<210> 897

<211> 143

<212> PRT

<213> Homo sapiens

<400> 897

Met	Tyr	Thr	Thr	Val	Asp	Ala	Asn	Gly	Tyr	Leu	Lys	Asn	Gly	Ser	Ala
1				5					10					15	
Gly	Gln	Leu	Ser	Gln	Ser	Ser	His	Leu	Ala	Leu	Gln	Leu	Pro	Tyr	Asn
			20					25					30		
Val	Leu	Gly	Leu	Gly	Arg	Ser	Ala	Asn	Phe	Leu	Asp	His	Leu	Tyr	Val
		35				40						45			
Gly	Ile	Pro	Arg	Pro	Ser	Gly	Glu	Lys	Ser	Ile	Arg	Lys	Gln	Glu	Trp
	50					55					60				
Thr	Ala	Ile	Ile	Pro	Asn	Ser	Gln	Leu	Ile	Val	Ile	Pro	Tyr	Pro	His
65					70					75					80
Asn	Val	Pro	Arg	Ser	Trp	Ser	Ala	Lys	Leu	Tyr	Leu	Thr	Pro	Ser	Asn
			85					90						95	
Ile	Val	Leu	Leu	Thr	Ala	Ile	Ala	Leu	Ile	Gly	Val	Cys	Val	Phe	Ile
		100						105					110		
Leu	Ala	Ile	Ile	Gly	Ile	Leu	His	Trp	Gln	Glu	Lys	Lys	Ala	Asp	Asp
		115					120					125			
Arg	Glu	Lys	Arg	Gln	Glu	Ala	His	Arg	Phe	His	Phe	Asp	Ala	Met	
		130					135					140			

<210> 898

<211> 68

<212> PRT

<213> Homo sapiens

<400> 898

Met Met Asp Thr Met Asp Pro Phe Trp Phe Asn Ala Phe Lys Arg Thr

448

```

1           5           10           15
Asn Thr Ile Leu His His Leu Arg Met Ser Lys His Thr Asp Ala Ala
                20           25           30
Glu Glu Val Leu Leu Glu Lys Lys Gly Cys Thr Gly Val Ile Thr Leu
                35           40           45
Asn Arg Pro Lys Phe Leu Asn Ala Leu Thr Leu Asn Met Ile Arg Gln
                50           55           60
Ile Tyr His Ser
65

```

<210> 899
 <211> 74
 <212> PRT
 <213> Homo sapiens

```

<400> 899
Met Ala Ser Ser Thr Val Pro Val Ser Ala Ala Gly Ser Ala Asn Glu
1           5           10           15
Thr Pro Glu Ile Pro Asp Asn Val Gly Asp Trp Leu Arg Gly Val Tyr
                20           25           30
Arg Phe Ala Thr Asp Arg Asn Asp Phe Arg Arg Asn Leu Ile Leu Asn
                35           40           45
Leu Gly Leu Phe Ala Ala Gly Val Trp Leu Ala Arg Asn Leu Ser Asp
                50           55           60
Ile Asp Leu Met Ala Pro Gln Pro Gly Val
65           70

```

<210> 900
 <211> 81
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> 19
 <223> Xaa = Arg,Ser

<220>
 <221> UNSURE
 <222> 75
 <223> Xaa = Leu,Arg

<220>
 <221> UNSURE
 <222> 74
 <223> Xaa = Lys,Asn

```

<400> 900
Met Gly Ala Thr Gly Pro Lys Leu Gly Ala Cys Gly Leu Arg Leu Gly
1           5           10           15
Gly Arg Xaa Gly Gly Ser Glu Ala Ala Gly Ala Glu Gln Ala Leu Val
                20           25           30
Arg Pro Arg Gly Arg Ala Val Pro Phe Val Phe Thr Arg Arg Gly
                35           40           45
Ser Met Phe Tyr Asp Glu Asp Gly Asp Leu Ala His Glu Phe Tyr Glu
                50           55           60
Glu Thr Ile Val Thr Lys Asn Gly Gln Xaa Xaa Ala Lys Leu Arg Ala
65           70           75           80
Ser

```

<210> 901
 <211> 68
 <212> PRT

<213> Homo sapiens

<400> 901

```

Met Trp Ala Phe Pro Trp Glu Glu Pro Leu Tyr Val Ile Thr Phe Tyr
1          5          10          15
Ile Val Ala Val Glu Ala Lys Trp Thr Glu Ser Pro Arg Gln His Glu
          20          25          30
Val Val Leu Gln Val Ile Leu Glu Lys Leu Thr Tyr His Thr Thr His
          35          40          45
Gln Pro Met Gln Lys Pro Val Tyr Val Gln Ser Ala Glu Cys Leu Gly
          50          55          60
Pro Pro Lys Lys
65

```

<210> 902

<211> 107

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 100

<223> Xaa = * ,Trp

<220>

<221> UNSURE

<222> 27

<223> Xaa = * ,Tyr

<220>

<221> UNSURE

<222> 23

<223> Xaa = Arg,Ser

<220>

<221> UNSURE

<222> 19

<223> Xaa = Glu,Gly

<220>

<221> UNSURE

<222> 74,107

<223> Xaa = Gly,Val

<220>

<221> UNSURE

<222> 102,104

<223> Xaa = His,Gln

<220>

<221> UNSURE

<222> 28

<223> Xaa = Lys,Thr

<400> 902

```

Met Pro Gln Leu Lys Leu Thr Ser Val Trp Thr Thr Thr His Ser Tyr
1          5          10          15
Tyr Thr Xaa Thr Arg Thr Xaa Ser Pro Gly Xaa Xaa Pro Gly Pro Pro
          20          25          30
Pro Met Asp Ile Gln Met Ala Asn Asn Phe Thr Pro Pro Ser Ala Thr
          35          40          45
Pro Gln Gly Asn Asp Cys Asp Leu Tyr Ala His His Ser Thr Ala Ser
          50          55          60

```

450

Ile Val Met Pro Leu His Tyr Ser Leu Xaa Phe Ile Ile Gly Leu Val
 65 70 75 80
 Gly Asn Leu Leu Ala Leu Val Val Ile Val Gln Asn Arg Lys Lys Ile
 85 90 95
 Ile Ser Pro Xaa Ala Xaa Pro Xaa Pro Ile Xaa
 100 105

<210> 903
 <211> 72
 <212> PRT
 <213> Homo sapiens

<400> 903
 Met Leu Asp Ser Ile Trp Lys Pro Asp Leu Phe Phe Ala Asn Glu Lys
 1 5 10 15
 Gly Ala Asn Phe His Glu Val Thr Thr Asp Asn Lys Leu Leu Arg Ile
 20 25 30
 Phe Lys Asn Gly Asn Val Leu Tyr Ser Ile Arg Leu Thr Leu Thr Leu
 35 40 45
 Ser Cys Pro Met Asp Leu Lys Asn Phe Pro Met Asp Val Gln Thr Cys
 50 55 60
 Ile Met Gln Leu Glu Ser Leu Val
 65 70

<210> 904
 <211> 57
 <212> PRT
 <213> Homo sapiens

<400> 904
 Met Gly Lys Gln Arg Asn Leu Thr Leu Phe Phe Ser Phe Pro Phe Cys
 1 5 10 15
 Glu Thr Val Phe Leu Asn His Leu Leu Asn Ser Val Phe Phe Pro Ser
 20 25 30
 Tyr Ser Phe Val Glu Pro Leu Leu Ser Tyr Leu Arg Phe Pro Gln Thr
 35 40 45
 His Ile Ser Leu Leu Phe Gln Leu Tyr
 50 55

<210> 905
 <211> 102
 <212> PRT
 <213> Homo sapiens

<400> 905
 Met Glu Arg Leu Thr Leu Pro Leu Gly Gly Ala Ala Ala Val Asp Glu
 1 5 10 15
 Tyr Leu Glu Tyr Arg Arg Ile Val Gly Glu Asp Asp Gly Gly Lys Leu
 20 25 30
 Phe Thr Pro Glu Glu Tyr Glu Glu Tyr Lys Arg Lys Val Leu Pro Leu
 35 40 45
 Arg Leu Gln Asn Arg Leu Phe Val Ser Trp Arg Ser Pro Thr Gly Met
 50 55 60
 Asp Cys Lys Leu Val Gly Pro Glu Thr Leu Cys Phe Cys Thr His Arg
 65 70 75 80
 Tyr Lys Gln His Lys Thr Asp Leu Glu Ala Ile Pro Gln Gln Cys Pro
 85 90 95
 Ile Asp Pro Pro Ala Lys
 100

<210> 906
 <211> 104
 <212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 64

<223> Xaa = His,Gln

<400> 906

```

Met Ser Thr Leu Leu Ile Ser Ser Ser Asp Ala Leu Leu Glu Glu Leu
1          5          10          15
Glu Arg Ser Thr Leu Gln Asp Ser Asp Glu Tyr Ser Asn Pro Ala Pro
          20          25          30
Leu Pro Leu Asp Gln His Ser Arg Lys Glu Thr Asn Leu Asp Glu Thr
          35          40          45
Ser Glu Ile Leu Ser Ile Gln Asp Asn Thr Ser Pro Leu Pro Ala Xaa
          50          55          60
Ser Cys Ile Leu Pro Ile Ser Arg Ser Ser Met Ser Thr Val Lys Pro
65          70          75          80
Lys Ser Gln Arg Asn His His Arg Leu Leu Lys Arg Gln Gln Leu Leu
          85          90          95
Ser Trp Met Ser Ser Leu Leu Thr
          100

```

<210> 907

<211> 63

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 57

<223> Xaa = Ala,Pro

<220>

<221> UNSURE

<222> 60

<223> Xaa = Gly,Val

<220>

<221> UNSURE

<222> 39

<223> Xaa = Phe,Val

<220>

<221> UNSURE

<222> 56

<223> Xaa = Pro,Ser

<400> 907

```

Met Asp Tyr Gly Tyr Val Gln Thr Thr Ser Thr Glu Met Leu Arg Asn
1          5          10          15
Phe Ile Gln Thr Glu Ala Val Val Ser Lys Pro Phe Ser Leu Phe Asp
          20          25          30
Leu Ser Ser Val Gly Leu Xaa Gly Ala Glu Thr Gln Gln Ser Lys Val
          35          40          45
Ala Pro Gln Gln Cys Ser Gln Xaa Xaa Arg Pro Xaa Gln Ser Leu
          50          55          60

```

<210> 908

<211> 65

<212> PRT

<213> Homo sapiens

<400> 908

```

Met Leu Ser Ala Leu Pro Gly Trp Gly Pro Ala His Leu Gln Arg Pro
1          5          10          15
Leu Leu Gly Pro Ala Ser Cys Leu Gly Ile Leu Arg Pro Ala Met Thr
          20          25          30
Ala His Ser Phe Ala Leu Pro Val Ile Ile Phe Thr Thr Phe Trp Gly
          35          40          45
Leu Val Gly Ile Ala Gly Pro Trp Phe Val Pro Lys Asp Pro Thr Ala
          50          55          60
Glu
65

```

<210> 909

<211> 83

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 42

<223> Xaa = Asp, Glu

<400> 909

```

Met Leu Val Glu Arg Lys Val Asn Gly Leu Val Pro Arg Leu Met Pro
1          5          10          15
Val Ile Ala Ala His Trp Glu Ala Gln Val Gly Ala Ser Leu Glu Gly
          20          25          30
Ser Ser Ile Leu Asp Arg Ala Val Ile Xaa His Asn Leu Leu Ser Ala
          35          40          45
Ser Lys Leu Tyr Asn Asn Ile Thr Phe Glu Glu Leu Gly Ala Leu Leu
          50          55          60
Glu Ile Pro Ala Ala Lys Ala Glu Lys Ile Ala Ser Gln Met Ile Thr
65          70          75          80
Glu Asp Val

```

<210> 910

<211> 94

<212> PRT

<213> Homo sapiens

<400> 910

```

Met Ala Ser Asn Glu Val Ile Ala Asp Ile Asn Cys Lys Gly Arg Ser
1          5          10          15
Lys Ser Asn Leu Gly Trp Thr Pro Leu His Leu Ala Cys Tyr Phe Gly
          20          25          30
His Arg Gln Val Val Gln Asp Leu Leu Lys Ala Gly Ala Glu Val Asn
          35          40          45
Val Leu Asn Asp Met Gly Asp Thr Pro Leu His Arg Ala Ala Phe Thr
          50          55          60
Gly Arg Lys Val Lys Ile Ile Leu Cys Ser Met Phe Val Ser Glu Val
65          70          75          80
Phe Gly Gly Val Val Thr Ile Val Phe Ser Val Ile Thr Ile
          85          90

```

<210> 911

<211> 75

<212> PRT

<213> Homo sapiens

<400> 911

```

Met Pro Gly Pro Thr Pro Ser Gly Thr Asn Val Gly Ser Ser Gly Arg
1          5          10          15
Ser Pro Ser Lys Ala Val Ala Ala Arg Ala Gly Asp Pro Leu Ser Gly

```

453

			20					25					30				
Arg	Gly	Lys	Met	Pro	Ala	Val	Gly	Gln	Gly	Val	Gln	Ala	Ala	Gln	Pro		
		35					40					45					
Arg	Gln	Ala	Pro	Gly	Gly	Cys	Gly	Asp	Ser	Thr	Gln	Lys	Ile	His	Leu		
	50					55					60						
Gly	Ser	Lys	Leu	Ala	Leu	Phe	Gln	Tyr	Trp	Leu							
65					70					75							

<210> 912

<211> 160

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> 107

<223> Xaa = * ,Trp

<220>

<221> UNSURE

<222> 97

<223> Xaa = Cys,Trp

<400> 912

Met	Arg	Pro	Leu	Pro	Val	Gly	Asp	Val	Arg	Leu	Glu	Leu	Ser	Pro			
1			5					10					15				
Pro	Pro	Pro	Leu	Pro	Val	Pro	Val	Val	Ser	Gly	Ser	Pro	Val	Gly			
			20				25					30					
Ser	Ser	Gly	Arg	Leu	Met	Ala	Ser	Ser	Ser	Ser	Leu	Val	Pro	Asp	Arg		
		35				40					45						
Leu	Arg	Leu	Pro	Leu	Cys	Phe	Leu	Gly	Val	Phe	Val	Cys	Tyr	Phe	Tyr		
	50				55					60							
Tyr	Gly	Ile	Leu	Gln	Glu	Lys	Ile	Thr	Arg	Gly	Lys	Tyr	Gly	Glu	Gly		
65				70					75					80			
Ala	Lys	Gln	Glu	Thr	Phe	Thr	Phe	Ala	Leu	Thr	Leu	Val	Phe	Ile	Gln		
			85					90					95				
Xaa	Val	Ile	Asn	Ala	Val	Phe	Ala	Lys	Ile	Xaa	Trp	Asp	Arg	Thr	Arg		
			100					105					110				
Ser	Trp	Leu	Tyr	Ala	Ala	Cys	Ser	Ile	Ser	Tyr	Leu	Gly	Ala	Met	Val		
		115				120						125					
Ser	Ser	Asn	Ser	Ala	Leu	Gln	Phe	Val	Asn	Tyr	Pro	Thr	Gln	Val	Leu		
		130				135					140						
Gly	Lys	Ser	Cys	Lys	Pro	Ile	Pro	Val	Met	Leu	Leu	Gly	Val	Thr	Ser		
145					150					155					160		

<210> 913

<211> 60

<212> PRT

<213> Homo sapiens

<400> 913

Met	Arg	Leu	Ser	Pro	Phe	His	Leu	Gln	Leu	Glu	Phe	His	Met	Lys	Glu		
1			5						10					15			
Lys	Arg	Glu	Asp	Leu	Gln	Ile	Ser	Trp	Ser	Phe	Ile	Ser	Val	Pro	Glu		
			20					25					30				
Met	Ala	Val	Asn	Ile	Gln	Pro	Lys	His	Trp	Gly	Arg	Thr	Arg	Trp	Leu		
		35				40					45						
Arg	Gln	Val	Arg	Cys	Leu	Thr	Phe	Ser	Arg	Thr	Ser						
	50					55					60						

<210> 914

<211> 137

<212> PRT

<213> Homo sapiens

<400> 914

```

Met Glu Tyr Leu Ile Gly Ile Gln Gly Pro Asp Tyr Val Leu Val Ala
1      5      10      15
Ser Asp Arg Val Ala Ala Ser Asn Ile Val Gln Met Lys Asp Asp His
20      25      30
Asp Lys Met Phe Lys Met Ser Glu Lys Ile Leu Leu Leu Cys Val Gly
35      40      45
Glu Ala Gly Asp Thr Val Gln Phe Ala Glu Tyr Ile Gln Lys Asn Val
50      55      60
Gln Leu Tyr Lys Met Arg Asn Gly Tyr Glu Leu Ser Pro Thr Ala Ala
65      70      75      80
Ala Asn Phe Thr Arg Arg Asn Leu Ala Asp Cys Leu Arg Ser Arg Thr
85      90      95
Pro Tyr His Val Asn Leu Leu Leu Ala Gly Tyr Asp Glu His Glu Gly
100     105     110
Pro Ala Leu Tyr Tyr Met Asp Tyr Leu Ala Ala Leu Ala Lys Ala Leu
115     120     125
Leu Gln Pro Thr Ala Met Val Pro Ser
130     135

```

<210> 915

<211> 55

<212> PRT

<213> Homo sapiens

<400> 915

```

Met Ala Asp Arg Gly Gly Val Gly Glu Ala Ala Ala Val Gly Ala Ser
1      5      10      15
Pro Ala Ser Val Pro Gly Leu Asn Pro Thr Leu Gly Trp Arg Glu Arg
20      25      30
Leu Arg Pro Gly Trp Arg Gly Leu Gly Pro Arg Cys Gly Ser Trp Arg
35      40      45
Gly Trp Gly Cys Phe Thr Pro
50      55

```

<210> 916

<211> 132

<212> PRT

<213> Homo sapiens

<400> 916

```

Met Ile Ser Ile Thr Glu Trp Gln Lys Ile Gly Val Gly Ile Thr Gly
1      5      10      15
Phe Gly Ile Phe Phe Ile Leu Phe Gly Thr Leu Leu Tyr Phe Asp Ser
20      25      30
Val Leu Leu Ala Phe Gly Asn Leu Leu Phe Leu Thr Gly Leu Ser Leu
35      40      45
Ile Ile Gly Leu Arg Lys Thr Phe Trp Phe Phe Phe Gln Arg His Lys
50      55      60
Leu Lys Gly Thr Ser Phe Leu Leu Gly Gly Val Val Ile Val Leu Leu
65      70      75      80
Arg Trp Pro Leu Leu Gly Met Phe Leu Glu Thr Tyr Gly Phe Phe Ser
85      90      95
Leu Phe Lys Gly Phe Phe Pro Val Ala Phe Gly Phe Leu Gly Asn Val
100     105     110
Cys Asn Ile Pro Phe Leu Gly Ala Leu Phe Arg Arg Leu Gln Gly Thr
115     120     125
Ser Ser Met Val
130

```

<210> 917

<211> 143
 <212> PRT
 <213> Homo sapiens

<400> 917

```

Met Leu Gln Ala Pro Arg Gly Arg Ala Arg Ala Ala Leu Arg Arg Leu
1      5      10      15
Ala Val Ala Thr Arg Thr Met Ala Gly Ala Pro Thr Val Ser Leu Pro
      20      25      30
Glu Leu Arg Ser Leu Leu Ala Ser Gly Arg Ala Arg Leu Phe Asp Val
      35      40      45
Arg Ser Arg Glu Glu Ala Ala Ala Gly Thr Ile Pro Gly Ala Leu Asn
      50      55      60
Ile Pro Val Ser Glu Leu Glu Ser Ala Leu Gln Met Glu Pro Ala Ala
65      70      75      80
Phe Gln Ala Leu Tyr Ser Ala Glu Lys Pro Lys Leu Glu Asp Glu His
      85      90      95
Leu Val Phe Phe Cys Gln Met Gly Lys Arg Gly Leu Gln Ala Arg Ser
      100     105     110
Trp Pro Gly Val Leu Asp Thr Leu Gly Leu Ala Thr Thr Leu Glu Pro
      115     120     125
Ile Glu Asn Gly Trp Arg Lys Arg Val Arg Gln Glu Ala Ala Tyr
      130     135     140

```

<210> 918
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 918

```

Met Lys Met Ser Arg Asn Ser Ser Leu Gly Gly Ala Thr His Cys Thr
1      5      10      15
Gly Ala Ser Trp Glu Ala Lys His Gly Pro Ala Ser Gly Ala Ser Pro
      20      25      30
Ser Ser Leu Gly Phe Arg Pro Leu Ser Pro Ser Leu Pro Ala Leu Gly
      35      40      45
Leu Lys Glu Ala Pro Asp Ser Met
      50      55

```


INTERNATIONAL SEARCH REPORT

In. lonal Application No

PCT/IB 01/00914

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C07K16/18 C12N15/63 C12N5/10
 A01K67/00 A61K31/7088 A61K38/16 C12Q1/68 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A01K A61K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

SEQUENCE SEARCH, MEDLINE, BIOSIS, EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 91 11515 A (UNIV SOUTH FLORIDA) 8 August 1991 (1991-08-08) 98.4% identical to SEQ ID NO:170 of present application. 98.8 % identical to SEQ ID NO:1 figure 1 -----	1-12, 14-19, 22-25
X	EP 1 067 182 A (HELIIX RES INST) 10 January 2001 (2001-01-10) the whole document -----	1-12, 14-19, 22-25
X	WO 97 39030 A (GENETICS INST) 23 October 1997 (1997-10-23) the whole document -----	1-12, 14-19, 22-25

☐ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- * & * document member of the same patent family

Date of the actual completion of the international search

26 March 2002

Date of mailing of the international search report

27. 05. 02

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

Herrmann, K

INTERNATIONAL SEARCH REPORT

international application No.
PCT/IB 01/00914

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 20 and 23 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 13, 20, 21, 26
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-19 and 21-25 all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: claims 1-19 and 21-25 all partially

Isolated polynucleotides comprising a nucleic acid sequence encoding the polypeptide of SEQ ID NO:170 such as a polynucleotide comprising a nucleic acid sequence as in SEQ ID NO:1 and subject-matter relating thereto. Polypeptide comprising an amino acid sequence as in SEQ ID NO:170 and subject-matter relating thereto.

Inventions 2-408: claims 1-19 and 21-25 all partially

Idem as subject 1 but limited to each of the polypeptides as in SEQ ID Nos:171-338, 456-560 and 785-918 and polynucleotides as in SEQ ID Nos:2-169, 339-455 and 561-784, respectively. Invention 2 is limited to subject-matter relating to SEQ ID NOs 2 and 171, invention 3 to SEQ ID NOs 3 and 172, etc.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 13, 20, 21, 26

Claim 13:

A polynucleotide may be clearly defined by an amino acid sequence. However, this is not the case vice versa. A practically infinite number of polypeptides can be deduced from a polynucleotide (6 reading frames, varying lengths). For this reason a meaningful search cannot be carried out for the subject-matter of independent claim 13 (Art. 6 PCT).

Claims 20, 21 and 26:

Compounds as such are not sufficiently defined by their mode of action. Therefore, claims 20, 21 and 26 have not been searched because a peptide that specifically binds to the polypeptide comprising the sequence as in SEQ ID NO:170 (claims 20 and 21) other than an antibody or a modulator of the polypeptide comprising the sequence as in SEQ ID NO:170 (claim 26), respectively, is neither disclosed nor supported within the terms of Art. 5 and 6 PCT, respectively (Art. 17(2)(a)(ii) PCT).

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 01/00914

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9111515	A	08-08-1991	AU 1484595 A	22-06-1995
			AU 6482898 A	25-06-1998
			AU 7172900 A	01-02-2001
			AU 7301091 A	21-08-1991
			CA 2075196 A1	02-08-1991
			EP 0513201 A1	19-11-1992
			JP 3159216 B2	23-04-2001
			JP 6500228 T	13-01-1994
			WO 9111515 A2	08-08-1991
			US 5804176 A	08-09-1998
			US 5849534 A	15-12-1998
<hr/>				
EP 1067182	A	10-01-2001	AU 5849300 A	30-01-2001
			AU 5849400 A	30-01-2001
			AU 5849500 A	30-01-2001
			AU 5850500 A	30-01-2001
			EP 1067182 A2	10-01-2001
			EP 1197554 A1	17-04-2002
			EP 1203816 A1	08-05-2002
			EP 1201754 A1	02-05-2002
			WO 0104312 A1	18-01-2001
			WO 0104299 A1	18-01-2001
			WO 0104300 A1	18-01-2001
			WO 0104301 A1	18-01-2001
			JP 2002017376 A	22-01-2002
<hr/>				
WO 9739030	A	23-10-1997	AU 2732997 A	07-11-1997
			AU 2734497 A	07-11-1997
			AU 2801697 A	07-11-1997
			EP 0914336 A2	12-05-1999
			EP 0914337 A2	12-05-1999
			EP 0912731 A2	06-05-1999
			JP 2001509004 T	10-07-2001
			JP 2001510323 T	31-07-2001
			JP 2002502234 T	22-01-2002
			US 2001016650 A1	23-08-2001
			WO 9739123 A2	23-10-1997
			WO 9739102 A2	23-10-1997
			WO 9739030 A2	23-10-1997
			US 6280739 B1	28-08-2001